EVALUATION OF PERFORMANCE IN ACADEMIC AND SCIENTIFIC INSTITUTIONS

Clyde Manwell and C. M. Ann Baker

Covert discrimination cannot exist without falsehood...Often the blame is placed on the victim...

You will cause the least trouble to the university if you can be persuaded to doubt your own worth and to believe that your opponents are being objective. The university may also rely on your fear of what will happen to you if you resist its authority. If, however, you do struggle, the university will put forth a powerful effort to get you to fight the case on its own grounds (usually the quality of your scholarship, occasionally the quality of your teaching).

Those who have the power to practice discrimination understand its mechanisms far better than do their victims. Many of us are burdened with a sense of inferiority; though we know better, we are nevertheless likely to believe authorities who assure us in a seemingly rational and objective or even benevolent tone that our work is mediocre. People who possess the critical vocabulary, who know the lingo, can always find the words and means to support a predetermined negative judgment; all that is needed is the will to make that judgment.

There are stock methods of creating such a negative assessment: discovering ''inconsistencies'' between parts of your work; distorting the argument; suggesting that the development lacks complexity; asserting that your work failed to include some other aspect of the subject.

When discrimination is being practiced, the publication of an article, even in a prestigious journal, is no longer a sufficient guarantee of the article's quality...

Marcia R. Lieberman[1]

INTRODUCTORY REMARKS: OBJECTIVES AND SUMMARY OF THE TOPICS SURVEYED

The primary purpose of this chapter is to help individuals defend themselves against inadequate or inaccurate allegations. Many fail to fight good cases because they do not know the Marquess of Queensberry's Rules — and how some administrators hit below the belt and get away with it. As the above quotation from Marcia Lieberman points out, the academic system can force the innocent victim to believe in his or her own guilt. The system, with its potential for bias in peer review at a number of levels (thesis examination, refereeing of manuscripts and research grant proposals, hiring or firing, promotion and other awards), destroys much individual creativity.[2] This destruction is done in an inefficient and hypocritical way: discrimination for irrelevant reasons is practised in the name of academic excellence. The dissident is sacrificed on the altar of administrative convenience.

This chapter needs to be longer than previous ones for three reasons: first, there is a wide range of honest opinions about what constitutes good performance in research or teaching. However, complications arise in interpreting the words used to justify some guilty verdicts.

Second, when the suppressors strike, the victim may not have the time or the resources to sift through the literature on evaluation of performance in order to prepare an adequate defence. As Professor R. D. Wright described so well in the preface to Eddy's book on the Orr case, the victim will find himself deserted by his fairweather academic friends.[3] The victim, even if a dues-paying member of his staff association, may find that staff association officials regard him as a threat to their own negotiations for improved pay or working conditions, negotiations carried out in a cosy manner with the administration. While there are a variety of options open in countering suppression generally, the options are narrowed when the suppressors attempt to effect dismissal. Especially where the administration seeks publicity so that sacking becomes a means to discredit the dissident, the victim may be presented with no viable alternative but to fight his case — and fight it on the administration's terms. We choose the words "viable alternative" deliberately, for we know of two cases where unfairly victimised academics committed suicide. It is impossible in our review to anticipate all possible allegations, or procedural tricks. Our case histories reveal a sample. Our aim here is to provide enough background material, and a varied enough range of topics, to allow those who are unfairly discriminated against to prepare their best case.

Third, several of the scholars who have attempted to improve the accuracy and fairness of the evaluation of performance have themselves been subjected to harassment. These case histories are included in this chapter. They provide additional evidence for the ubiquity of suppression in the self-styled "free world" of Western science and academe.

The organisation of this chapter, together with certain important conclusions, is summarised in five stages: first, there are five short sections providing background: why complaints centre on work performance rather than the dissident act itself; how the private ideology of academic power-brokers charges certain words and phrases with new and subtle meanings (words and phrases that often appear in letters of recommendation, dismissal complaints etc.); how "commitment" to research or to teaching, often an important criterion in hiring or firing, is situationally variable; how "personality", also a stock word in evaluating candidates, seems to be associated with teaching, research, dissidence and excellence — and how *radical chic* behaviour complements administrative *repressive tolerance*.

Second, we consider the *inputs*: the time and resources available to the individual whose performance is being evaluated. It is unfair to consider only large publication lists as a criterion for excellence without also considering such matters as the money available to buy equipment and to hire technicians — or, as covered in the chapter "Academic Exploitation", without also considering how some academics appropriate the efforts of students, staff and wives.

Third, we consider the *outputs*. These are often estimated by the opinions (frequently secret opinions) of other scholars. How reliable is *peer review*, especially with the absolute power enjoyed in secret peer review? This has many consequences, not least of which is the necessity to insist upon proper legal procedures in formal disciplinary actions. There must be examination and cross-examination of all witnesses before the defendant, who must receive a verbatim copy of all that is said or written.

Fourth, we consider the *outputs* as measured by the more quantitative methods of bibliometrics. This is much more than just counting publications. In recent years *Science Citation Index* has been used (and abused) in evaluating the performance of individuals. Properly used, the citation indices can be a protection against erroneous allegations, as in the example provided in the chapter "IMVS versus John Coulter: Charges and Cross-examination", charge number three. While there are many caveats, the citation indices are some protection from certain types of erroneous charges; in particular, they allow a comparison to be made of individuals who are, and who are not, up for administrative action. The antagonistic attitudes expressed by certain establishment figures suggest that they regard these bibliometric tools as a threat to their own "wheeling and dealing" in the reward system of science and academe.

Last, but by no means least, we survey means for the evaluation of teaching, probably the most controversial of all aspects of staff performance.

Why Complaints about Dissidents are Phrased in Terms of Unsatisfactory Performance — and Not the Dissident Act Itself

> ... the President of the University was a man for whom I conceived, I think justly, a profound aversion. If a lecturer said anything that was too liberal, it was discovered that the lecturer in question did his work badly, and he was dismissed.
>
> Bertrand Russell[4]

As mentioned in the Introduction to this book, the belief system which has become enshrined as "academic freedom" means that suppression is not usually considered openly justifiable by university administrations. The situation described in the chapter "The Fruit Fly Papers", where the first official letter of complaint about the performance of a senior staff member is solely an unparticularised charge that the senior staff member (and his wife) had made some public criticisms about a local pesticide-spraying program (which had affected them at their residence), would probably be considered too direct an infringement on academic freedom, if not freedom of speech, at most universities.

In general, complaints about the actions of dissidents in universities are phrased in terms of a failure to perform satisfactorily, usually in regard to their responsibilities for research or teaching.

Even in government research institutions, which make no pretence at "academic freedom", this rule is sometimes followed. Although in government laboratories sanctions against certain types of public comment are formalised (for example Official Secrets Acts, or their equivalent), the tendency is to phrase the complaint against the dissident more directly in terms of work performance. For example, there is the recent dismissal of Dr Ross Hesketh from the research laboratory of the Central Electricity Generating Board (UK), allegedly for failing to heed "proper management instructions".[5] It appears that Hesketh's real offence was to co-author a letter to the editor of *The Guardian* (with Professor Martin Ryle of Cambridge University), which suggested that British civilian nuclear reactors were generating weapons grade plutonium which was exported to the United States and used for nuclear weapons.

Why Complaints are Imprecise — Public versus Private Ideologies

In attempting to understand how criteria are chosen for describing performance in teaching and research, one must bear in mind the distinction between attitudes and behaviour. What academics say should be done, and what academics actually do, need not necessarily be the same. One example, epitomised in the slogan "publish or perish", is discussed later.

Attitudes may be expressed differently for different audiences. Superficially, we recognise two levels. Each set of attitudes expresses an ideology — but there is one for public display and another for consenting academics in private. The *private ideology* shows itself in many of the expressions used to describe performance. Marcia Lieberman's quotation[6] referred to "stock methods of creating... a negative assessment". Caplow and McGee's[7] classic study of the academic profession provides many examples in their quotations from interviews. Commonly used phrases are: "X lacks commitment", "Y has an unsuitable personality", and "Z may be bright but is he *sound*?"

These phrases are a kind of code, the expression of a private belief system. Like all ideologies it is promulgated by a power bloc for the purpose of preserving its privileges.

"Commitment" is to the hierarchical system itself, or to narrow academic specialisation. An academic who is said to lack "commitment" may have, in fact, shown too much commitment to unpopular ideas or social causes. Such an individual is regarded as a threat to academics who find it profitable to serve the state or business. The allegation of a lack of "commitment" is sometimes used when it would be better to make a straightforward allegation of inadequate research or publication.

"Unsuitable personality" can be translated as "likely to rock the boat". Allegations of "unsuitable personality" have two marked advantages over other types of allegations. First, it is so clearly a matter of opinion that it is virtually impossible to disprove. Second, it is an allegation that is likely to get a sympathetic hearing from many academics, who, in turn, have suffered from the actions of superiors whose personalities were "difficult", to use another word commonly found in letters of recommendation.

"Sound" as a description of research often means the routine application of accepted paradigms.

There are also stock words and phrases for where a power figure is pushing a protégé for a job or a promotion. For example, when confronted with a lack of publications, one can always write "he shows promise".

It can be difficult to penetrate this private ideology, with its subtle changes in the meanings of words and phrases. These subtle changes colour the way complaints about performance are expressed. These changes allow discrimination on irrelevant or marginal criteria to sound impressive.

The *public ideology* of academics is epitomised by the rhetoric at graduation ceremonies, as well as at other official occasions. Speakers use a combination of the time-honoured phrases of pompous elegance and the current trendy buzz-words: "dedication to truth", "without fear or favour", "producing the well-rounded man", and "the pursuit of excellence". The existence of case histories of the kind that fill this book is evidence enough that the public ideology is now little more than public relations window-dressing. It is regrettable that the disparity between expressed attitudes and actual behaviour has created so much cynicism. The true goals of educational institutions are being eroded and devalued.

The differences between public and private ideologies show in the way complaints are phrased. The public ideology allows appeal for a general concern about teaching, or "gross moral turpitude" (which, revealingly, is almost always used for cases of sexual misbehaviour rather than academic dishonesty, although the term could as well apply to plagiarism or to bringing false charges). A head of a school refers to the stock phrases in certain letters of recommendation[8]: "I do not trust this man with girls" — or the more damning "I do not trust this man with boys".

In summary of this section, in studying dismissal complaints, letters of recommendation and other forms of peer review, we note that allegations are often imprecise. The existence of conflicting public and private ideologies accounts for some of this vagueness. There is also the "Vicar of Bray" syndrome shown by many intellectuals: a desire not to commit oneself too far and, above all, to be on the winning side. It often becomes necessary to devote considerable effort to get complaints about performance properly particularised. Although not without its own difficulties in terms of professional bias and a complex vocabulary, legal advice can be useful in cutting through the woolliness in complaints about performance, and, in particular, getting to the kernel of specific, testable allegations.

Teaching versus Research?

There are strongly held views about whether the relationships between teaching and research are antagonistic, neutral, or mutually supportive. These views often influence the choice of candidates for jobs or for promotions. On occasion, these views are used to mask discrimination against dissidents.

Some universities and private colleges in the United States, and certain Oxbridge colleges in the United Kingdom, discriminate against researchers in their hiring policies. This is justified publicly on the grounds that research-motivated individuals are "too narrow" or "likely to neglect their students". One wonders if this policy, while often sincerely believed, has not arisen from a fundamental fear of dissidence. Good research sometimes poses a difficult challenge to authority. The kinds of higher educational institutions which profess this preference for teaching over research often either train the offspring of the rich or have a strong religious base.

Certain elite universities, notably Harvard in the US and Cambridge in the UK, have placed emphasis on research quality as well as teaching. Especially in the era after the Soviets launched Sputnik I, in 1957, the emphasis at a number of American universities shifted dramatically towards excellence in research. The quality of research was sometimes measured by quite crude criteria, as in the following advertisement for a vacant position: "Cellular immunologist wanted for position of associate professor level...Candidates must have documented proof of scientific quality, for example at least $100,000 per year in grant funds".[9] The sudden influx of large amounts of research money was accommodated by a corresponding ideological shift in the relative merits of teaching versus research. A belief developed that good researchers were automatically good teachers.

Despite the dogma and the money, there are few useful data to answer the question about whether teaching and research help or hinder each other. The situation also brings us back to that elusive word, so frequently used (or abused) by academics: personality.

In studying the staff of Canadian universities who were involved in the teaching of psychology, J. P. Rushton and colleagues[10] found some differences in the *average* for certain dimensions of personality when comparing good teachers with good researchers — but there was great individual variation in each category. The personality traits of the good teacher are largely neither congruent with, nor opposite to, those of the good researcher. Largely, the personality characteristics of good researchers and good teachers are orthogonal. However, certain characteristics are common both to good teachers and to good researchers, for example high intelligence and ability in leadership. The better researchers tended to be ambitious, aggressive, authoritarian, persistent, and non-supportive. The better teachers tended to be liberal, sociable, non-anxious, supportive, non-authoritarian, and aesthetically sensitive. But, what is especially important is the wide variance in each category. This means that the superficial assessment of personality by appointment committees is unlikely to be effective in the accurate assessment of the effectiveness of individuals.

Furthermore, even the thorough study by Rushton and his colleagues is limited to but a single subject discipline and to a homogeneous type of academic setting. It may well be that quite different personality traits will make for the best research, or best teaching, in other subjects, or in smaller, more democratically organised institutions. A few scientists have written perceptively of the personality traits required for research in different fields, for example the botanist Edgar Anderson[11], the physiologist Hans Seyle[12], and the molecular biologist Atuhiro Sibatani.[13] Their impressions have been confirmed in the results on psychological testing of scholars from different fields, although considerable variability occurs.[14]

Thus, it would be unwise to use casual estimates of personality traits to make decisions on hiring or firing in the absence of information on actual performance in teaching or research. Generalisations about the personality traits that preadapt individuals for teaching or for research in particular subjects must recognise the nature of the academic reward system, the vague vocabulary arising from public and private ideologies, the absence of precise data on psychological testing of the individuals being considered, and the rather random application of what criteria for good teaching, or good research, actually exist. When an administration places all the rewards on research, rather than a balance between research and teaching, it is hardly surprising that teaching and research become antagonistic. And, it can require a strong personality to get something done in spite of the system.

SCHOLARSHIP, DISSIDENCE AND RADICALISM

The personality of teachers and researchers can be examined in the perspective of conservative versus liberal attitudes. Good scholarship requires intellectual independence. There must be a constant willingness to challenge the status quo, including one's own preconceptions. This must be balanced by certain social restraints: a willingness to consider views differing from one's own; politeness to opponents; and, a refusal to let rank, rather than content, settle an argument.

In our case histories the dissidents usually had at least an average level of performance (when compared with their appropriate peers, matched for age, rank etc.). Yet, administrators had alleged that the performance of these dissidents was so bad as to warrant the rarely used, ultimate sanction: sacking. Information from the literature suggests that, although there are exceptions, dissidence (both social and intellectual) and performance are positively associated.

In the extensive survey of the social characteristics of university staff in the UK, A. H. Halsey and M. A. Trow[15] collected data for self-assessed general political and social characterisation, i.e., "far-left", "left", "centre", or "right". These characterisations were compared with the quality of their degrees, with the following results:

Self-reported assessment	*Number of Staff with particular classes of degrees* (I + IIa)	(III + IV + pass)	*Academic "quality ratio": (I + IIa) divided by* (III + IV + pass)
"far left"	8	2	4.0
"left"	105	83	1.26
"centre"	54	50	1.08
"right"	33	66	0.50

Treated as a 2 × 4 contingency table: $\chi^2 = 17.33$, *three degrees of freedom;* $p = 0.0006$

These results suggest that there is some positive association between self-reported radicalism and the quality of the degree which they obtained (in many cases a number of years before the interview). One might like more direct measures of dissidence versus performance. However, the quality of an academic's first degree is positively correlated with later performance.

It can also be argued that more objective criteria of the inclination towards dissent are needed than just self-reported assessment of "far left", "left", "centre" or "right". However, the tendency for brighter students and staff to be more radical has been observed in several American studies.[16] Radicalism in one direction can be offset by conservatism in other directions, quite apart from the inevitable arguments about what constitutes radicalism or conservatism. Clark Kerr has written: "Few institutions are so conservative as the universities about their own affairs while their members are so liberal about the affairs of others".[17]

Despite the reservations, there is sufficient tacit agreement among academics that dissidence and intellectual achievement are positively correlated that there has been the evolution of a phenomenon sometimes called *radical chic*: a small number of academics cultivate radical or eccentric behaviour. Unfortunately, their concept of radicalism is sufficiently distorted as to suggest that this includes rudeness, anti-intellectualism, unkemptness, and extremes of sexual behaviour — depicted so well in the British Broadcasting

Corporation's televised version of Malcolm Bradbury's *The History Man*.

This radical chic behaviour is usually tolerated by administrators. Sometimes it is rewarded (as in *The History Man* example, with the "hero" ultimately shown in his true reactionary colours). Such tolerance is a form of *repressive tolerance*: "Within the solid framework of preestablished inequality and power, tolerance is practiced indeed. Even outrageous opinions are expressed, outrageous incidents are televised."[18] Radical chic behaviour is welcomed because it allows administrators to prove they defend academic freedom. It does not threaten the existing power structure. It is not to be confused with dissidence.

While we would not wish to claim that all dissidents are excellent scholars, the data available suggest an association between, on one hand, academic and social dissidence, and, on the other hand, certain measures of ability in performance in the academic milieu. Thus, by penalising dissidents, the university selects against the very excellence that, in commencement day rhetoric, it claims to defend.

Suppression of Attempts to Improve the Evaluation of Research Performance: "Converging Partial Indicators"

There is a fascinating contradiction in the behaviour of some establishment scientists. On one hand, they show a willingness to rank order the accomplishments and abilities of other scientists. On the other hand, these same establishment scientists show an unwillingness to have their own accomplishments evaluated by relatively objective means.

This contradiction came to the surface recently when part of the English astronomy elite overreacted to a study on the scientific productivity of the Isaac Newton telescope and of Jodrell Bank. One reads: "legal proceedings, initiated by members of the Advisory Board for the Research Councils"[19] and "[Sir Bernard] Lovell threatened to take legal action against the journal *Research Policy* if it published . . . "[20]

The two researchers who precipitated this reaction, John Irvine and Benjamin Martin, of the Science Policy Research Unit of the University of Sussex, have provided a detailed study of methods for evaluating research performance. They have devised a method of "converging partial indicators", using several distinct criteria of performance, ranging from opinions from a large sample of relevant scientists, to the actual citation impact of publications.[21]

Irvine and Martin are hesitant to apply such methods in comparing individuals — even though some of the techniques are widely (if often inadequately) used by scientific administrators, for example taking opinions on what one scientist thinks about another scientist's research or teaching. Irvine and Martin prefer to apply their "converging partial indicators" to comparisons of different laboratories, departments or institutes, that is, to *groups* of individuals. This allows statistical comparisons to be made with much more discriminatory power. Nevertheless, on occasion it is necessary to evaluate the performance of individuals; in such instances the work of Irvine and Martin provides much good cautionary advice.

Inputs and Outputs: Measures of Opportunity and Quality?

It is a curious reflection on administrative values that the measurement of performance is so frequently incomplete. Despite the current emphasis on input/output economics, administrators often fail to consider all the input and output variables when assessing staff performance.

Too often, output is measured only in terms of *publications*, such as papers, reviews and books. However, less formal means of publication are occasionally of value, for example computer programs, bibliographic lists, printed notes, and laboratory direction sheets. All of these items, often circulated within "invisible colleges", can be important contributions to the research of others. Patents are a form of publication, emphasised more in technological institutions.

The object is often to estimate the *quality* of output. Few administrators we have encountered have read enough about the history of science (or other subjects) to realise just how difficult it is to judge quality. One indirect method is by *recognition*, the way the scholarly community expresses its collective judgement about the value of the contributions of an individual scholar. Of course, there are many examples of where proper recognition came only posthumously. Recognition can be measured by counting citations, a technique discussed later. Other measures of recognition include invitations to speak at meetings, to write reviews, or to assess manuscripts or research grant proposals. Recognition is also manifested by requests for advice from other scholars, or requests for assistance in building apparatus, and requests for help in locating source materials. Translation of articles from foreign languages is a valuable service, sometimes ignored in assessing scholarly performance. Recognition is often measured in terms of prizes or election to prestigious societies, but the reward systems of scholarship are not always accurate measures.[22]

There are two other types of output which may count positively or negatively, depending on the situation: participation in community organisations and communication in the mass media. In both situations, which sometimes overlap, an academic can play an important role, both in expanding knowledge and in disseminating it. However, whether or not such activities are counted for or against an academic can on some occasions depend upon whether or not administrators approve of the organisations or the nature of the communication.

Participation in the "wrong" community organisations can result in peer disapproval, or even administrative action. Participation in the affairs of business or government is usually regarded as laudatory, although sometimes it can result in neglect of students and research, or in serious conflict-of-interest problems.

Attitudes about communication in the mass media present some contradictions. On one hand, academics have often looked down on popularisation. However, there is in the hiring process a desire to appoint individuals whose names are recognised, not only by other scholars but by influential members of the wider community.[23] Thus, visible scholars are sometimes sought after.

The sheer size of the scholarly community, plus the marked increase recently in the number of unemployed and underemployed intellectuals, all competing for very limited job opportunities, has resulted in some academics using communication in the mass media. Furthermore, this is sometimes encouraged by administrators who are attempting to influence politicians in order to reverse the trend to budget cutting. Many universities have employed media relations officers in order to feed information to local newspapers and television. An increasing number of academics realise that they must sell their product to the tax-paying public. But at the same time, this has meant that university administrations are becoming more sensitive about the occasional academic who criticises publicly some vested interest, especially if it can be tapped for money.

There still is, even if in an eroded form now, a general academic ethos of avoiding publicity. This ethos serves business and government, for it gives access to important information for entrepreneurs and bureaucrats but denies it to the less powerful.[24]

We would argue that, given the definition of a university as a centre for the discovery and dissemination of knowledge, the less formal means of scholarly communication should not be ignored in evaluating performance. At least the standards should be clearly spelled out. There should not be one rule for corporate interests and another rule for community interests.

Inputs, especially time

An inconsistency in the evaluation of performance is a failure to take into account differential access to *inputs*. The question should be asked in comparing staff members: what opportunities have the individuals been given — or denied?

Allowance must be made for *professional age*, the number of years spent in research or teaching. It is accepted that time spent in the armed services, especially in time of war, is exempted. There were opinions voiced at the recent Australian tenure hearings, conducted by Senator Teague's committee[25], that allowance should be made for women to take time off for child-bearing. It was even suggested that this activity should be counted towards professional advancement. We would suggest that this is a purely personal matter — and we note that no mention was made of the problems of many single or married women who spend considerable time caring for parents or other relatives.

Allowance must be made for how duties are distributed in the organisation. Since not all staff in a university are equally capable or motivated in teaching and research, it is reasonable that duties be distributed with some allowance for this. However, it is unacceptable that an individual, assigned years of heavy teaching loads, should be expected to have the same count of publications as colleagues who have had light teaching loads. In a university teaching and research are interdependent activities. There should not be a wide variance in the time that different staff members have for these twin responsibilities.

Administration

A disturbing trend, visible in some Australian universities, is to elevate administration to the status of a duty comparable with research or teaching. Indeed, Sir Mark Oliphant complained about the tendency of Australian universities to reward administration more than either teaching or research.[26] Since Sir Mark made his complaint, the situation has worsened. In the last five years of relatively declining budgets the University of Adelaide lost over 60 teaching staff members. Almost none have been replaced — but a nearly identical number of administrative positions have been filled. We find it disturbing that universities can maintain constant numbers of administrative staff, or even increase numbers, while the number of teaching staff declines — and the number of new doctorates, who are unable to find a job in academe or research, continues to rise.

Why do administrators put more emphasis on outputs than inputs in evaluating staff?

In many cases examined by us we have noted that administrators put much more emphasis on outputs rather than inputs. The fact that a staff member may have been denied fair shares of research funds, or given an extra heavy teaching load, is ignored when publications are counted. We suggest that the answer to this question is simple: administrators have much more *control* over inputs.

This is not to deny the possibilities that an establishment can influence access to publication outlets — for example the massive amount of financial support for publication which the CIA administers clandestinely to willing academic recipients.[27] However, local elites do have more control over staff members' access to time for research and to research funds, whereas local elites usually cannot influence decisions concerning the rejection or acceptance of manuscripts submitted for publication in international journals.

Many research granting agencies require, or at least allow, a secret statement on the applicant and his intended research to be made, often by a department head. This occurs in spite of the fact that in many cases the administrator is clearly outside the area of his professional competence to pass on the merits of the researcher or his proposal. Abuse of this administrative prerogative was admitted by a former Chairman of the Australian Research Grants Committee:

> In the early years of its operation, the [Australian Research Grants] Committee asked as a matter of routine for a detailed comment and recommendation by Heads of Department. It has ceased to ask in this way for these reports, leaving it to the discretion of the Head of Department to make a comment where he considers it is needed. The simple situation was that most Heads of Department appeared to have only swans on their staff. Where *Y*, a Head of Department, reported "*X* is a goose", the Committee was inclined to wonder what *X* and *Y* had been fighting about.[28]

It is with such opportunities to keep dissidents from getting a fair share of inputs that an administrator can, almost effortlessly, make the allegation of inadequate performance a self-fulfilling prophecy.

In some cases administrators, or other staff members, intent on harassing a dissident, have interfered in research by removing equipment (that did not belong to them) or by taking away technical staff. This happened to Dr Struan Sutherland at the Commonwealth Serum Laboratories in what has been called "the bureaucratic civil war of ludicrous pettiness".[29] We had similar experiences at the University of Adelaide. Furthermore, Clyde Manwell was assigned additional teaching duties while preparing his defence against the dismissal charges. The fact that these duties were the same as what the complainant felt were unsatisfactory suggests that the administration either didn't care about the duties or didn't really believe the complaint.

Thus, as administrations have so much power in controlling access to inputs, it is not too surprising that, in comparing the performance of staff members, they pretend that everyone has had an equal chance. Hence, more emphasis is placed on comparing outputs, rather than inputs.[30]

Opinions of Other Scholars: Peer Review in Hiring, Firing, Publication and Research Grants

> We took him on the basis of the enthusiastic support of an outstanding professor at Harvard. That's very important. If Princeton pushes a man, I know it means I'll have to look somewhere else. I don't trust Columbia either, or Chicago. With one or two exceptions in each department, those bastards are shysters; they'll say anything about anyone to get a man placed.

From one of the interviews conducted by
Theodore Caplow and Reece J. McGee[31]

In the attempt to measure *quality* of output, *peer review* forms the basis for the commonest method. It is also commonly abused. In several dismissal cases the administration justified its action largely or solely on the basis of one or two adverse opinions, solicited from individuals who had a clear conflict of interest or a known personal dislike for the individual being assessed.

A distinguished expert in computer science and artifical intelligence has commented as follows on the peer review sampling procedures that are occasionally used:

> A fairly senior official of a [research granting] agency, whose nationality I shall not disclose, once told me that if for any reason he felt justified in short-circuiting the system in order to get a given result, he would make a judicious selection of referees — either the scientist's particular friends or his particular enemies, according to which result he wanted. He needn't have told me. I knew it already. In his heart of hearts so does any scientist who has been in the game any length of time.[32]

There are examples that suggest that errors and bias in peer review result in the frequent rejection of good manuscripts. One editor rejected at least three manuscripts which, when ultimately published elsewhere, helped to win their authors Nobel Prizes. Hans Krebs' first description of the citric acid cycle, Urey's work on heavy hydrogen, and Fermi's research on beta decay were all rejected by *Nature*.[33] Other examples of inconsistencies and differences in opinion have been observed in studies on peer review.[34] Among the most conspicuous failures of peer review are the examples of the publication of faked or plagiarised research.[35] Another type of failure of peer review is where conspicuous errors are published; scholars in many fields have their favourite examples.

Errors in the use of statistics are common — although there are strong differences in opinion about the most appropriate statistical tests or the possible inferences in certain cases.[36] Three statisticians who examined a highly prestigious medical publication concluded:

> Sixty-two reports that appeared as Papers and Originals . . . in 13 consecutive issues of the *British Medical Journal* included statistical analysis. Thirty-two had statistical errors of one kind or another; in 18 fairly serious faults were discovered. The summaries of five reports made some claim that was unsupportable on re-examination of the data.[37]

IMPORTANCE OF OPEN COURT PROCEDURES IN EXAMINING OPINIONS

An episode related to a case history is useful here. It is chosen because it illustrates several problems dissidents have encountered. The example shows the importance of the careful evaluation of opinions, in particular the use of the methods evolved by the legal profession for examination and cross-examination of witnesses in open court.

A dismissal complaint included an allegation of "a number of flagrant mistakes" in the use of the Chi-squared statistic in a book (which was not about statistics but had used some statistical procedures in one chapter). The complainant wrote that these mistakes "will undoubtedly have upset the general conclusions of the book but how profoundly I do not know". The complainant then made a request that the Professor of Genetics be asked to comment.

In correspondence to the Vice-Chancellor the dissident admitted one statistical error (in the number of degrees of freedom in a Chi-squared test) and pointed out that another "error" was simply a typesetting mistake which had no other implications. It was, however, maintained that two other alleged errors were themselves incorrect. It was also maintained that, even if all the alleged errors were indeed errors, these were a very small part of the book and not important to its purpose or its conclusions.

This was apparently not acceptable to the administrative powers that be. In the ensuing correspondence the complainant reasserted his views and now stated that the Professor of Genetics was in agreement with him.

Later the contents of the letter of complaint were examined in open court. The Professor of Genetics was *not* listed to appear in court as a witness to support the claim (by the complainant) that he had agreed with the selection of four errors. Surely, had he been that certain of the importance of those errors, he would have been willing to appear.

In court the complainant, under cross-examination, admitted that the alleged errors were not important — and that some of the errors were not as erroneous as he had thought at the time. The upshot of all this was to leave a residue, which basically featured the one mistake admitted from the very beginning. Even that mistake did not alter the specific conclusion, let alone other aspects of the book.

In contrast, another professor, knowledgeable of statistical techniques (and Fellow of the Academy of Sciences) was willing to appear for the dissident as an expert witness, although the results of the cross-examination, combined with a settlement before judgement, made further examination unnecessary.

It can be argued strongly that such matters should not have to be settled in court. However, when a university administration acts as if such allegations are important (and accepts a dismissal complaint with such allegations when the Statutes allow the Vice-

Chancellor to refuse to accept the complaint), fairness demands that its contents be evaluated critically.

The episode shows that opinions in secret correspondence can differ significantly from opinions expressed in proper, formal, non-secret procedures. Complaints about such details in a dissident's work are common. This is not to excuse error but to maintain a sensible perspective in evaluation of an individual's work. Secret opinions are open to suspicion.

PROPER PROCEDURES FOR OBTAINING OPINIONS

"Honesty is the best policy." Any solicitation for a professional assessment should state clearly the reasons for the solicitation. The individual being assessed should receive copies of all correspondence. Secrecy is out.[38] If an individual's job is at stake, proper court procedures for the examination and cross-examination of witnesses should be followed. Care must be taken to ensure that the sample of opinions is made fairly and that the sample size is reasonable. One possibility is to allow each side to select an equal number of names as possible assessors. In this way biases for and against will tend to cancel out. Care should be taken to avoid soliciting opinions from individuals who have a conflict of interest. This is often easier said than done, given the nature of the academic profession. The person who is being assessed should be able to challenge opinions that come from individuals with a clear conflict of interests, including personal dislikes.

We would also insist on the *control experiment*: a similar solicitation of opinions about the peers of any individual who is being considered for sacking. After all, the question is whether or not an individual is doing such a bad job, in comparison with his colleagues who are not being considered for dismissal, as to warrant the ultimate sanction. Peers should be of two types: colleagues in his department, or related departments within the institution, of approximately similar professional age and rank; and his "invisible college" of colleagues at other institutions throughout the world who are doing research in the same subject.

Inside or outside the institution?

A point of some contention is the extent to which opinions should be obtained from individuals inside or outside the university or research institution where an individual's performance is being evaluated. On some matters only individuals in the same institution can make knowledgeable comment, but in reference to professional accomplishments the comments are best obtained from individuals located at other places. Outside opinion is not always as independent as one might wish. The outside assessor may be a direct competitor, or a friend of some important party involved in the complaint.

Probably the safest situation is where a large number of opinions are randomly sampled. One can argue endlessly about the fairness of certain procedures. It is the blatant disregard for natural justice that should be opposed. In one case the individual who authored a dismissal complaint was allowed to assist in choosing two academic jurors, whom he later acknowledged in writing were in agreement with him. Moreover, this information was kept from the individual being considered for sacking. We doubt if many individuals would wish to be tried under such a "kangaroo court" system.

Limitations of letters of recommendation

It is chastening to read the only detailed study we know of on letters of recommendation.[37] Scientists and non-scientists alike frequently placed more emphasis on trivial and irrelevant factors in letters of recommendation than information on research or teaching. Comments on female candidates are particularly prone to be focused on physical attractiveness, on qualities which are perceived by the writer as feminine or masculine, and on speculations about commitment to research or to child-bearing.

Clyde Manwell has seen several hundred letters of recommendation and finds no reason to disagree with Lionel S. Lewis's conclusions.[40] In truth, it is extremely difficult to predict how well an individual will perform. As the very nature of letters of recommendation will determine the candidate's chances, such letters are likely to be self-fulfilling prophecies.

The historical perspective on opinions

In evaluating opinions, administrators who wish to do a fair job, or dissidents who wish to protect themselves, should read the literature on peer review.[41, 42, 43] It is useful to examine historical cases from science or other forms of scholarship. Had Mendel been considered for dismissal from the Abbey in Brno, there would have been no difficulty in finding experts who considered his research to be worthless, or worse.[44] Mendel was so far ahead of his time that it took nearly forty years before biologists began to realise that Mendel had, almost singlehandedly, revolutionised the science of genetics, and thus much of the biology that depends upon it. Furthermore, R. A. Fisher's suggestion, that Mendel (or his assistant) "doctored" his experimental results, has been countered by several researchers.

SUPPRESSION OF ATTEMPTS TO STUDY PEER REVIEW

The most original approach to studying peer review is that performed by Douglas P. Peters and Stephen J. Ceci.[45] They took twelve papers which had been written by prominent psychologists and published in high status journals. The papers were typed out as new manuscripts. The texts were left essentially unchanged. However, Peters and Ceci replaced the prominent names, and also the prominent institutional affiliations, with imaginary names and institutional affiliations. In other words, the papers were left unaltered except that the authorship and institutional affiliation were changed from high status to low status. What then was the fate of these manuscripts which were, in fact, almost verbatim copies of papers already considered by peer review to be good enough to be published?

Of the 38 referees and editors involved, only three detected the resubmitted articles. Of the nine manuscripts that proceeded further into the peer review process, eight of the nine were rejected. In many instances the grounds for rejection were "serious methodological flaws"!

Behavioral and Brain Sciences published Peters and Ceci's paper, together with the comments of 55 scientists from a wide variety of disciplines, including many individuals with experience in refereeing papers and several individuals who were editors of journals. The comments, and the cited references, provide a wide range of opinion about peer review. Some, though not all, editors and other establishment figures objected to the Peters and Ceci study, largely on one or both of two grounds: the sampling procedure itself, in particular the small sample size, and the ethics of deception. We do not believe that these two grounds are valid criticisms of the Peters and Ceci study. Indeed, we do not believe that these two grounds, taken together, represent an entirely consistent point of view. If the experiment were unethical, would it not be compounding the unethicality by taking a larger sample, or by variations in the procedures which would amount to further deception of referees and editors? It is like telling a child: What you did was wrong — but you should have done more of it.

Being a referee or an editor is a sacred position of trust. Especially with so much peer review being done in secret, those who are in the extremely powerful position that secrecy automatically confers must be held accountable. Peters and Ceci executed a simple experiment which established a fact that many suspected but others denied. Their demonstration, that peer review is at best a chancy process, is of the greatest importance. Not only is the morale, indeed the livelihood, of many researchers dependent upon peer review in getting jobs, promotion, papers published, or receiving research grants, but peer review also serves society as a whole, for we are all dependent upon the flow of accurate information. Even results in scholarly publications become translated into action by politicians, by bureaucrats and by industrialists. Such individuals often practise "the principle of unnatural selection"[46], choosing from among the variety of opinions and contradictory "facts" those items which are perceived as most useful for the actions they wish to rationalise to a sceptical public. Peer review, in its broadest sense (which includes the publication of criticism of already published articles), reduces the opportunities for error and for antisocial action.

The research by Peters and Ceci represents dissidence at its best. Their demonstration of the imperfections in peer review, as currently practised by establishment figures, constitutes a healthy challenge to their power. It is instructive to read in Ceci and Peters' own words what happened to them:

> Upon collection of the data we entered a period lasting approximately two years during which we experienced an intense and negative reaction from many powerful individuals in our profession for having conducted our study. One editor in the study wrote a letter threatening a lawsuit for copyright violations. Actually, we had obtained permission from the original authors and copyright holders (publishers) to use their materials in our study.
>
> Quite unexpectedly, several editors who had not been directly involved with our study wrote scathing letters calling into question our professional ethics because of our use of deception (which, according to our national code of ethics, requires a careful cost-benefit analysis before employing). Actually we had given serious consideration to alternative, nondeceptive means of examining peer review but we ended up rejecting them. There simply was no experimentally sound way to study the issues we were interested in studying without using some form of deception. [We agree fully with Ceci and Peters.]
>
> The field of psychology has a long, somewhat tarnished, history of using deception in studies where the subjects are relatively powerless individuals, like children, college students, and unsuspecting citizens in the community. In our investigation the subjects were anything but powerless. We felt that if ever deception was justifiable, it was in the context of our study because the data were of great potential importance to the academic and research community and there were no sound, nondeceptive means of collecting them. Given the large role of peer review in determining people's lives (e.g., tenure, promotion, hiring, grant awards, professional status and recognition), it is urgent that we learn more about the practice and specifically more about factors influencing a reviewer's judgement before we assign [disproportionate] importance to that judgement in, say, a promotion decision.
>
> Recognizing the potential hazards, such as embarrassment to editors, we never divulged names of editors or journals...
>
> Other negative repercussions included several threats to professionally censure us, and threats to reject the work of our colleagues, supposedly because they had been part of a department that had approved such ethically bankrupt research.
>
> This charge is sadly ironic. Our department had no screening mechanism for nonfunded research, thus they were neither asked nor had they given, their approval, although two senior colleagues were consulted about the study's design. Later, when our chairman was personally contacted by an angry editor, he withdrew all of our departmental resources until we finished the study. We were sent a memorandum informing us that typists, photocopy, mail, etc. were "off limits" to us as long as we continued using the procedures we had adopted. When challenged on the grounds that this amounted to a violation of our academic freedom [the department chairman] offered to have the decision reviewed by the entire faculty. Because of the timing (right in the middle of a tenure review), we declined his offer and moved the project off campus. It was completed at our own expense.
>
> These personal attacks took their toll... Finally, after two unsuccessful attempts to publish our findings, replete with personally insulting, *ad hominem* reviews, we found a publisher and positive reviews. Soon press releases were telling a diverse audience of our findings. Letters of support (over a thousand) came pouring in. Every one was complimentary.[47]

Publication Counts: the Mythology of "Publish or Perish"

Given the uncertainties in personal opinions, some administrators have sought more quantifiable means for comparing candidates in regard to research productivity. It is common among academics to make jokes about administrators counting numbers and pages of publications, or even weighing them. It has all become part of the mythology of "publish or perish".

Yet, providing it is not the only criterion which is used, and providing it is done in a sensible way, such publication counts do have some value. Among other things they prove that the obsessive concern about "publish or perish" is unwarranted by evidence from the *realpolitik* of academic institutions.[48] Lionel S. Lewis, a sociologist of the academic profession, has concluded:

> ...the idea that one must publish or perish is somewhat of a myth, perpetuated by the notoriety given a few cases. But it is no accident that [this myth] is widely accepted. First, it is a protection for those who wish to rid themselves of an unwanted colleague; it is a real explanation to conceal the true reasons for some dismissals. Second, it helps [to keep] those who are fearful of losing their positions busy doing research and writing (a good deal of which may be useless), diverting them from an active role... Third, it promotes the idea that an objective standard is used in arriving at decisions which are really made subjectively.
>
> Although academicians quarrel with the dictum of publish-or-perish, there appears to be little interest in investigating its credibility. Nor is there an attempt to abolish it; this might lead to an awareness that its importance has been exaggerated. It is too convenient a myth to abandon.[49]

The fact is that a researcher who is conscientiously performing his scholarly duties will be doing *some* publication. He may be prolific and write lots of little papers with minimal information content — "the least publishable unit".[50] He may be a perfectionist and write only a few papers in a lifetime. In the latter case, however, those papers are likely to represent either major breakthroughs or large reviews or presentations of new data, works of monographic size. Thus, counting the pages as well as the number of papers reduces the discrimination against the perfectionist. The use of citation index counts is also sometimes a protection for the perfectionist, for his smaller total output is likely to get proportionately greater recognition, that is, a higher citation count per paper per unit time; but care must be taken to allow for the large difference in citation counts between different subjects.

Another justification for counting pages as well as publication numbers is as follows. Although exceptions abound, on the whole editors guard journal space. Editors often insist that authors remove material from otherwise acceptable papers. Thus, the opportunities for padding are fairly limited. The variance may be great, but on average larger papers represent more work.

Some would argue too that allowance must be made for the format of different journals. A one-page paper in a journal like *Science* or *Nature*, with three columns of fine print per page, is likely to contain more information per page than a ten-page paper in a single-column format and with large print.

There are important differences in publication style, both between subjects and between individuals in a given subject. Our impression is that in certain sciences there is relatively little variation in the size and structure of papers, for example mathematics, physics and chemistry, but in biology and in many non-scientific disciplines there are great differences in publication styles. For example, biologists who specialise in the classification of animals and plants (taxonomists) have diverse publication styles. Many taxonomists write lots of short papers, describing one or a few species at a time, or reporting an extension of the range of a previously described species. Other taxonomists prefer the perfectionist approach; their lifetime output is a few big monographs, but each monograph deals with hundreds of species. The different styles each have advantages and disadvantages. Somewhat similar individual variation is common in the arts and humanities. Certain scholars prefer to publish a number of book-length studies, others favour many short articles.

On the whole there is some association between quantity and quality[51,52] but that association is not so strong that one should become obsessive about it. We all know of examples of where an individual leaves only a few papers, but where those publications are of seminal importance for many other researchers.

Access to publication differs markedly between subjects. A far greater proportion of manuscripts submitted to social science journals are rejected than those submitted to physical science journals.[53] Contrary to the situation for most other fields of study, medical journals receive lucrative subsidisation from advertisements placed by "ethical" drug companies. This has allowed the proliferation of journals publishing in biomedical fields. Has this resulted in any deterioration in the standards of published work in biomedicine? On the other hand, biologists often find it very difficult to find publication outlets for large studies, unless they happen to be lucky enough to be at a museum or university which has its own "house journal". This situation does not always select for high quality because such "house journals" often have lax peer review procedures, or no peer review at all.

This is not the place for a detailed discussion of publication counts but three complications need to be considered: abstracts, multiple authorship, and journal prestige or rank.

Abstracts: the Quantum of Publication?

In some subjects a common form of publication is the *abstract*, a short summary of work in progress. Abstracts are often published as summaries of research presented at conferences. Certain scholars have perfected the art of padding their publication lists with abstracts and other non-definitive means of communication. First, there is the "paper read at conference"; this is followed by an "abstract", often with an impressive title nearly as long as the abstract itself; there may even be a separate "paper published in the proceedings of the conference" (and in some cases, to be fair, that may be a definitive paper); finally, hopefully, there is the definitive publication in one of the standard journals. With some care in the wording of titles and the arrangement of authors (in multiple-author works) there can be several "papers read at conference" and "abstracts" for each definitive paper that is finally published.

One way of detecting abstracts in publication lists is their shortness, usually a page or less in length. Occasionally job applicants evade this detection tactic by providing only the first page for *all* their publications.

Abuse of abstracts is probably greatest in biomedical research, which is highly competitive and has absorbed some of the ethics of the medical profession and the drug companies. However, we have seen also the use of abstracts to pad out publication lists in other subjects.

A large proportion of all the abstracts in biomedicine represent research which is never published in a definitive form.[54] There are many possible reasons for this, and some of the reasons are no fault of the authors. However, it is not unknown for the non-appearance of a definitive paper to be the consequence of the authors' inability to repeat the original experiment. This problem of the lack of reproducibility has resulted in a not inconsiderable amount of scientific humour, with jokes about journals with mythical names like "*The Journal of Unreproducible Results*" or "*Acta Retracta*".

Another problem of abstracts is *self-plagiarism*, where the same work is repeated in two or more different abstracts. This problem is not unknown in definitive publications, but is rarer there. Abstracts are rarely refereed before publication; this state of affairs may encourage more self-plagiarism.

This is not to criticise the use of abstracts in scientific communication. Abstracts play a role in establishing contact between different researchers. For those concerned about priority (and the reward system of science forces that concern on many scientists), abstracts have some significance. Abstracts form a convenient way to place small amounts of important information into the scientific literature in a way that is economical of journal space.

Our object here is to ensure that abstracts are not used to inflate publication counts and thereby not used to set unreasonably high expectations as to the number of published items. Certain conscientious and capable researchers only rarely use abstracts as a means of communication. Furthermore, the opportunity to publish (and to pad bibliographies with) "abstracts", "paper read at conference", and similar non-definitive publications, depends considerably on financial support, either from the institution or from research grants. Universities can be very capricious in providing funds and free time for staff to attend conferences and thus accrue an impressive list of abstracts and conference-related publications. We have a number of examples of where Australian universities have discriminated against dissidents in this way, not allowing them to attend, or refusing to contribute to expenses, even when the invitations came from such organisations as the Royal Society, the US National Academy of Sciences, the Israeli Academy of Science and the academy of sciences of various eastern European countries.

Multiple Authorship: Who Did the Work?

Probably the most vexatious question in evaluating research performance through publications is: who did the work? As science tackles the more complex problems, which often demand the use of a number of difficult and specialised techniques, more and more research involves groups of people in collaboration. In the period 1960 to 1980 the average number of authors on each paper increased from 1.67 to 2.58 (see note 50). In physics, especially high energy physics, occasional papers have 20 to 50 co-authors.

If allowance is not taken of multiple authorship in comparing peers, there is the likelihood of penalising the perfectionist, for most of his work is solitary, or done with one or two other researchers. Large teams of researchers can be expected to publish more papers per unit time.[55]

Whether or not a name is included in multiple authorship does not depend entirely on whether or not the individual made a substantial contribution to the research effect. Non-Ph.D. subordinates, and women generally, whatever their degree status, are frequently excluded from co-authorship, even though they have done much of the work.[56]

This problem is especially serious for students. Some staff feel that they have a kind of *droit de seigneur*, a right to appropriate the first research efforts of their students. In fairness, on some occasions the staff member has put a lot of time and effort into a rescue operation. Also, an occasional staff member will lean over backwards and include a student's name when it is not warranted — but more likely that type of action will be done when it is the name of a famous individual or a powerful academic administrator. Postdoctoral fellows usually get their contributions recognised by co-authorship, even if there are conflicts about name ordering. It is realised that success in recruiting more postdoctoral fellows into a large laboratory depends upon placing the former fellows into good jobs. To this end they must receive some share of co-authorship. However, even junior staff are not safe from having their efforts appropriated by senior "operators".

It is also difficult to assign relative credit on the basis of the position of names in multiple authorship. The usual rule is that the individual who did the largest share of the work is first author. Groups where two or more individuals had essentially equal roles often rotate the authorship positions in a series of papers.

Occasionally, alphabetical ordering of names is adopted, either by certain laboratories or (very rarely) as an official editorial policy. Harriet Zuckerman found differences in the etiquette of name-ordering, both between subjects and even between different workers in the same subjects. Her data for Nobel Prize winners are instructive and she concluded:

> On the assumption that authors' names are listed in order of the value of their contributions, [Nobel] laureates should be first-authors more often than other scientists; in fact, they are not. Instead, they exercise their *noblesse oblige* by giving credit to less eminent co-workers increasingly as their eminence grows. They do so more often after the prize...[57]

A further complication is that, with the increasing necessity to obtain grant funds for research, the successful "operator" who is adept at getting money insists upon co-authorship, even though he has played little if any role at the laboratory bench or in the analysis of the data. There is also a tendency for a group to wish to include the names of eminent individuals, or powerful scientific administrators, both to curry favour and to convey status on the group's efforts. William Broad quotes the Managing Editor of the *American Journal of Psychiatry*:

> ...it's sometimes a glory kind of thing, putting the chairman in (as coauthor) even though he was not directly involved, trying to bask in the light of a greater name.[58]

Broad also describes one attempt to solve the question of how much effort each co-author contributed. Each co-author was asked independently what percentage of a particular paper represented his contributions. One would hope that the sum of these estimates for all co-authors in a given publication would not be far from 100 per cent. In fact, the totals were as high as 300 per cent!

Given all the uncertainties over co-authorship, we believe that simple correction factors should be applied. One possibility is to divide by the number of co-authors. Another possibility is to give somewhat more weight to the first position, for example $2/(n+1)$, counting the other efforts all equally as $1/(n+1)$. For a useful discussion of various weighting systems see the review by West, Hore and Boon.[59]

Journal Prestige: "It's not only what you publish but where you publish"

Quite commonly administrators guess the quality of individual papers from their general assessment about the quality of the journals in which they are published, although, as our opening quotation[60] suggests, this behaviour can be modified by the circumstances. There is a strong competitive drive for some scholars to publish in what they consider to be high prestige journals. There is also sometimes a kind of "halo effect", where an individual defines a journal as "high quality" if *he* publishes in it, or is on the editorial board.

Access to a high prestige journal sometimes depends upon being the protégé of a powerful patron. Providing you have a good friend in the Royal Society or the National Academy of Sciences, you can escape both the rigours and the biases of the usual peer review process. This is a delicate issue, seldom discussed openly. An exception is Stephen Fretwell's[61] description of how the now widely recognised ecologist Robert MacArthur had great difficulty in getting his earlier papers published. It was only because he found a distinguished patron, G. Evelyn Hutchinson, that he could bypass peer reviewing by other ecologists and get his papers into the *Proceedings of the National Academy of Sciences, U.S.A.* Fretwell concludes his discussion of this sensitive topic by asking: "Without the Proceedings of the NAS, I wonder how far MacArthur would have gotten?" Although no one can object to the publication of the seminal papers by MacArthur, there have been so many examples of abuse by the bypassing of peer reviewing that members of the US National Academy of Sciences have complained, in a few cases publicly[62], and some reforms have been recently instituted.

Thus, many good quality papers get published in low prestige journals and vice versa. Some of the most serious faking scandals in science, and also cases of plagiarism, have appeared in high-ranking journals.[63, 64] Nevertheless, *some* weight should be given to where a candidate has published, providing allowance is made for the complicating factors. There is a useful rank ordering of journals by citation impact.[65] It has certain limitations but is the only way to get away from the subjective judgements of individual administrators. It is imperative to bear in mind that acceptance or rejection of a manuscript is rather capricious. The system is noisy, with many biases, including bias against dissidence, both scholarly and social.

The words of the editors of the very prestigious *Physical Review Letters* bring us back to the peer review theme:

> It is a plausible hypothesis that, in spite of the responsible and conscientiously prepared counsel that we receive from our referees and Divisional Associate Editors, our selection of papers is too often ruled by chance. Under ordinary circumstances, each newly received paper is sent to two referees at once. *The agreement between the referees is scarcely better than chance.* A simplistic model which states that one-sixth of the papers are so clearly superior that each referee approves and one-sixth are so clearly poor that each referee advises rejection, leaving the other two-thirds to be decided by the flip of a coin, fits the correlations that we observe. On this model, *if two-thirds of the papers that we accept were replaced by two-thirds of the papers that we reject, the quality of the journal would not be changed.*[66] [Emphasis added]

Also, the words of Lynn Margulis, given below, turn our attention back to peer review when we consider judging individual contributions solely on the basis of where they are published. Margulis is a fine example of the dissident researcher. She originally championed the theory of the origin of subcellular organelles from separate microorganisms (the endosymbiont theory) against a unified disbelieving establishment. Now, as a result of research by her, and by many others, this theory is widely accepted, at least for chloroplasts and mitochondria. Margulis writes of her own experiences with peer review:

> Every one of our papers containing new ideas was rejected at least once; this was especially true of a prize-winning paper that generated 1,100 reprint requests...
>
> My experience on the question of a possible endosymbiotic origin for the microtubule system is a case in point. Six years ago I was told by an NSF [National Science Foundation] grants officer... that "important" scientists did not like the theory presented in a book I had written and that they would never fund my work. I was actually told that I should never apply again to the cell biology group at NSF, since my work "only appealed to the small minds in biology... the naturalists".[67]

CITATION INDICES: CATCHING "THE EEL OF SCIENCE"

> How Index-learning turns no student pale,
> Yet holds the Eel of science by the Tail.[68]

That refrain from Alexander Pope's *Dunciad* is part of his testimonial for a candidate for the Chair of Dullness. The lines, published in the eighteenth century, seem strangely appropriate in summarising the utilisation in the 1960s and 1970s of citation indices.

In publications scholars cite the work of other scholars, as well as of themselves. These citations can be used to construct a measure of information flow. The study of citation patterns has been used to work out the linkages in different fields of study.[69]

Highly cited papers or books are important landmarks in a subject. Citation indices, thus, can also be used to measure the impact of publishing scholars, although there are many caveats discussed below. Between the original *Science Citation Index*, starting in 1961, and the later *Social Science Citation Index* and *Arts and Humanities Citation Index*, essentially all forms of academic scholarship are now covered.

The basic procedure is to look up the individual's name in the appropriate citation index. One counts the number of times the publications by that individual are cited. One can also look up the items which do the citing in order to find out what type of use is made of the particular scholar's work.

At this point one runs into a problem. By 1978, *Science Citation Index* sampled over 2500 journals, including nearly 495,000 authored source items, which, in turn, had cited slightly over 7,475,000 items. *Science Citation Index* in 1978 refers to 920,039 "unique cited authors"

— but, in fact, these authors are not "unique" in the sense of being individually identified. Individuals with the same surname and the same initials, for example all A. Smiths, are lumped together. These "homographs" can be a source of confusion. (Why bibliometricians do not call these homonyms, the more usual word for identical names, instead of homographs, we do not know.) When in doubt, each cited item must be checked. Obviously, homographs are most likely to occur for common surname and initial combinations, but there are examples even for very unusual names.

Problems in the Interpretation of Citation Index Counts

Given the potentialities of citation counts for both use and abuse, it is desirable to have a brief discussion of some of the complications:

1. *Professional age*

The longer an individual has been publishing, the greater the opportunity to be cited. Nevertheless, although there is some controversy about the matter, with the possibility of different patterns in subjects differing in coherence and in the rate of growth, the professional age of the author of a cited item does not have much effect on the number of citations it receives, at least in sociology.[70]

2. *Differences between and within fields of study*

Citation practices vary greatly between disciplines, journals and authors. For science as a whole the average publishing author who is cited at all receives close to eight citations per year, though the actual figure is probably closer to seven because of the "homograph" confusion mentioned earlier. In the social sciences the average number of citations is about four. On the whole, chemists cite twice as many items per paper as do zoologists; thus, a chemist should get roughly twice as many citations as a zoologist of comparable age, quality etc.

3. *The time factor: differences in persistence*

Related to items 1 and 2 above is the time factor, which is partially dependent upon the differences in the rates of growth in different fields. In a faster growing field, relatively more papers become obsolete quickly. For science as a whole, the average cited paper has a "half-life" of only five years. Furthermore, in various fields between one-third and one-half of all papers are not cited again (or, to be more accurate, not cited again in the journals covered by the citation indices, a point to which we shall return). In biochemistry 62 per cent of all references were within the last five years of the date on the publication in which they were cited, whereas in sociology only 40 per cent were within the last five years.[71] Our impression is that researchers in unpopular areas, for example the classification of esoteric groups of animals or plants, get few citations in the short term; however, their work is likely to be relatively persistent, also picking up a few citations many years later.

4. *Reasons for citing a paper*

There are many reasons for citing a paper and useful short summaries of these reasons have been prepared by Lawani[72] and Gilbert.[73] Most of these reasons are, however, favourable to the use of citation counts as an approximate measure of impact. Nevertheless, the author's choice of citations is often consciously or unconsciously chosen so as to persuade the reader, and this results in subtle bias.

It is commonly argued that citation counts are meaningless because of the "fact" that many citations are made for the purpose of calling attention to errors. In fact, it is relatively rare for a paper to be cited only because of a trivial error in it. A paper may occasionally be cited because the citer disagrees with the cited, but that does not necessarily mean the citer is right and the cited is wrong. Even when it does, one must remember that science progresses in part by falsification (to use a Popperian oversimplification). Citing a work because the citer has, with the benefits of time and the further advances of science generally, a better technique or a wider perspective, is all part of scholarly progress. Needless to say, the best way to avoid being cited for errors is to publish nothing at all.

5. *"The Public Relations Web"*

This phrase has been used by Rodger Mitchell[74] and Leigh Van Valen[75] to denote the situation where the work of a particular ingroup is promulgated by confining citation and other forms of recognition to within the ingroup, failing to cite outsiders' publications, even when they have clear priority. We believe that this is probably the most important bias in using the citation indices, either to evaluate individuals or to trace out the origins of scientific developments.

Using an example provided by Daryl Chubin and a colleague, Timothy Lenoir[76] points out that the individual who had most clearly postulated the existence of "reverse transcriptase" (the enzyme that transfers information from RNA to DNA, an RNA-dependent DNA polymerase) had the importance of his work "severely underrepresented by citation counts". Lenoir concludes that "... biomedical researchers at large laboratories tend to project similar citation profiles due to intra-laboratory co-authorship and self-citation. The effect is to inflate the importance of work done at these large laboratories".

In the prolonged and intense rivalry between Andrew Schally and Roger Guillemin, in pursuit of the chemical structure of the hypothalamic releasing factors which control the output of hormones from the pituitary gland, Nicholas Wade refers to the allegations of "... the practice of citing as little as possible of each other's work".[77]

The noted historian of science, Derek J. de Solla Price, who had much to do with demonstrating the utility of quantitative methods, including citation counts, warned of "... the evident malpractice of some authors in preferentially citing their own papers, those of their special friends, or those of important scientists that confer status on their work".[78]

Comments about the failure to cite the relevant research of others are often heard in gossip among scientists. No doubt some of these comments are exaggerated. It is easy to assume the elitist mode and dismiss such comments as "the squeals of second-raters", as one scientist said to us. It is, therefore, salutary to read the words of Sir Gustav Nossal, an eminent immunologist and Director of what many consider to be the top medical research institute in Australia:

> The process of citation, which used accurately ... guide[s] readers to the sources that inspired a particular piece of research, is now frequently nothing more than a kind of game, and a rather dirty one at that. In its most extreme form, the game seeks to hide the foundations for the work amidst a mass of trivial or irrelevant references, and seeks to establish the author's own laboratory as the sole source of wisdom... In a complex competitive world, these problems will not disappear. Rather, they will intensify...[79]

6. *Citation as ritual*

In certain subjects very important work may not be cited, as the source is regarded as common knowledge. On the other hand, a few older references become *de rigueur* as citation landmarks, for example Darwin's *Origin of Species*. Here too, however, the "public relations web" has left its imprint. It is generally accepted that Alfred Russel Wallace independently discovered the theory of evolution by natural selection and published the outline in the same year as Darwin. Furthermore, it is known that there were important predecessors who had formulated the essential elements of the theory.[80] In contrast to Alfred Russel Wallace, Darwin came from a wealthy and intellectually prominent family. Perhaps it is no surprising, then, that he received most of the immediate recognition, with the result that hi name provided the convenient mark of recognition for subsequent citation. True, Darwir provided in *The Origin of Species* the most exhaustive documentation for the theory of evolution by natural selection (though in spite of the title Darwin devoted very little of that book to the central problem of evolution, the origin of species). On the other hand, Wallace's contributions in scholarly analysis were by no means unsubstantial. The present organisation of zoogeography owes much to Wallace — and to another neglected pioneer, Alfred Wegener, whose ideas on "wandering continents" were considered lunacy or heresy by the geological establishment for half a century.

It is also common for early important work to be missed. Although Mendel received a few citations prior to his "rediscovery"[81], it was only after the beginning of the twentieth century that citations revealed his true impact. Thus, there are time-dependent fashions in science, with Kuhn's[82] periods of paradigm change, or "revolutionary science", alternating with periods of gradual growth, or stasis. The periods of paradigm change will be marked by changes in ritual citation.

There is an aspect of ritual citation which is in need of study. We had noticed, in tracing out the literature on animal domestication, that a number of authors cited Hahn's *Die Haustiere und ihre Beziehungen zur Wirtschaft des Menschen*, usually for the first suggestion of religious and aesthetic factors as motives for animal domestication. Yet, we have been unable to locate a copy of this book in either the United Kingdom or Australia, even with the aid of interlibrary loan services. If the book is that difficult to obtain, one wonders how many of those who have cited this work have actually read it.

Somewhat similar is the pattern of ritual citation observed by Erwin Chargaff:

> ...bibliographies usually are wafted in their entirety from one paper to the next, except for the insertion of the respective authors' own contributions which, if luck has it, may then accompany, plasmidlike, the standard package in its subsequent passages.[83]

Biases in the Citation Index Structure Itself — and the Possibility of Scholars Changing Their Citation and Publication Behaviour

The citation indices only tabulate citations to the *first* author of a publication. A researcher whose name appears only in the second or subsequent positions in multi-authored papers will have a citation index score of zero. Will this intensify further the struggle for the first authorship position on papers?

An analysis of authors contributing articles to the *Journal of Physiology* is pertinent here. That journal has had the unusual practice of listing the co-authors in alphabetical order. The result is that, in comparison with other physiological journals, where the authors can be in any order, the *Journal of Physiology* has relatively more authors whose names begin with A–E and relatively fewer authors whose names begin with P–Z.[84] Thus, the publication behaviour (in this case, choice of journal for publication) can be influenced by authorship position.

Another serious bias in the citation index structure is that, until recently, *Science Citation Index* did not record the citations in the bibliographies of books, an extremely important nodal point in communication. The citation indices are still haphazard about picking up citations in books which are actually collections of papers or review articles. The problem is that the reference citations may be scattered through the text (usually as footnotes, though not invariably), at the ends of chapters, or collected together at the end of the book.

In general, the structure of the citation indices reflects the American dominance of many (but by no means all) areas of scholarship, dominance in terms of numbers, not necessarily quality. We have found that many important eastern European publications are missed. Although *Science Citation Index* has been used to compare the impact of scientists from different countries[85], we wonder how much these results will be biased by differences in the availability of publications, differences in the ease with which scientists can read different foreign languages, and the increasing use of the citation indices as bibliographic tools.

Biases in the Sources Sampled by Citation Indices and Current Contents — Information Capitalism versus Left-wing Journals?

The Institute for Scientific Information, which brings out the various citation indices, also brings out *Current Contents*, with different modules for different major research areas. *Current Contents* is essentially made up of copies of the table of contents of the most recent issue of a number of journals (often with an editorial by Eugene Garfield, the founder and director of the Institute of Scientific Information). The selection of journals for *Current Contents* is said to be on the basis of the frequency with which these journals are cited.

Jon Wiener[86] alleges that *Current Contents* is biased against left-leaning journals. For a time the appropriate modules of *Current Contents* did not include *Monthly Review*, a socialist American journal of considerable importance.

Not being listed in *Current Contents* probably has a serious effect on visibility. There are so many journals that scholars have great difficulty in keeping up, even in the narrowest specialties. Many researchers use *Current Contents* as a quick way to browse through the contents of the journals which are of most interest to them. *Current Contents* is extremely useful, for it allows researchers to know quickly about the existence of articles in journals not carried by their libraries, an important service in Australia where the combination of woefully inadequate journal subscriptions and the delay, from one to six months (as most journals are sent by sea mail to Australia), aggravates the already serious intellectual isolation.

The convenience of having the listing of titles and the authors' addresses in *Current Contents* means that many researchers (or their students or secretaries) flip through the listings and send off reprint request forms to the authors, or obtain a copy through their library. The decision to select certain source items for further inspection is reached largely on the basis of the title of the paper, although no doubt the reputation of the authors and the institutions from which the research was published also have some importance. A biochemist at Cambridge University, E. F. Hartree[87, 88], published a modification of a commonly used method for determining the amount of protein in samples. Because of a minor difference in the way Hartree's address was printed in the journal itself and in *Current Contents* he was able to score how many reprint requesters had seen the article in the journal in which it was published versus how many reprint requesters had learned of it from *Current Contents*. There were 375 who saw the actual article — compared with 2125 who used *Current Contents*.

This example warns us that *Current Contents* does play a dominant role in determining *access* to a researcher's publication. If the Institute for Scientific Information decides that a journal is not important enough to be included in *Current Contents*, many researchers (perhaps a majority, as in the above example) will miss relevant papers. While some of these may be picked up later by other literature retrieval methods, for example abstracting journals or the citation indices, the situation has all the elements of self-fulfilling prophecy. Many libraries now use *Current Contents*, or the citation index journal rankings, as criteria for deciding whether or not to purchase certain journal subscriptions.

Given that new journals will not have any citation score, and thus start at the bottom rank, they would automatically be excluded from *Current Contents*. The alternative is for the Institute for Scientific Information to favour some new journals which it *thinks* will be important.

All of this places an inordinate amount of power in the hands of one organisation — a privately owned organisation, not accountable to society, or its elected representatives, nor to the scientific community. While it can be said that, as a private corporation, the Institute for Scientific Information is accountable to the marketplace, it has no competitors and therefore completely controls the marketplace. The prices charged for many of its publications are so high that they are beyond the means of nearly all scientists and some libraries.

The Institute for Scientific Information is very profitable — and with its "fleet of chauffeur-driven cars [which] includes a Cadillac, a Lincoln, [and] a Jaguar"[89], it has clearly become indulgent in the luxuries of monopoly capital. Perhaps then, we can see a little better Jon Wiener's[90] concern about the absence of left-wing journals. George Orwell would have understood.

However, to be fair, the inventor of the citation abstracts, Eugene Garfield, who is also President and Chairman of the Board of the Institute for Scientific Information, and owns 65 per cent of its stock[91], has tempered his enthusiasm for his product with a warning to the scientific community:

> Like most other scientific discoveries, this tool can be used wisely or abused. It is now up to the scientific community to prevent abuse of the SCI [*Science Citation Index*] by devoting the necessary attention to its proper and judicious exploitation.[92]

We are now faced with an ominous situation; not only is there potential for error in the evaluation of individuals and groups of individuals, but there is also the potential for ultimately distorting the development of many fields of scholarship. The fault for allowing the Institute for Scientific Information to be a private organisation rests with the bureaucrats and scientific administrators in the US National Science Foundation, who failed to follow up Garfield's promising initial efforts in creating the citation index system and who refused to set up a publicly owned information agency.[93] There is really no excuse. The opportunity arose in that immediate post-Sputnik era, when funds were readily available and the Soviets had developed an impressive information service for their own scientists.

The result of that faulty decision within the US National Science Foundation is likely to have far-reaching consequences: first, as mentioned earlier, the high prices charged by the Institute for Scientific Information already limit the access to these useful bibliographic tools. Libraries must pay high charges for the ISI's subscriptions. This has come at an unfortunate time, for the economic recession has resulted in drastic cuts to the budgets of many university libraries, which have responded by purchasing fewer books and by cancelling journal subscriptions.

Second, leaving the situation in private hands may have inhibited further improvements. For example, probably only government financing and direction from the scholarly community could accomplish the necessary access to totally computerised citation indices that would so greatly speed the search for useful references, as well as assist in the analysis of the structure of science and other forms of scholarly activity. One vital project is to extend the citation indices back in time so as to be able to obtain a better historical perspective on the emergence of the structure of science.

Third, ways need to be found to prevent the system from being perverted by self-fulfilling prophecy. Especially with the citation indices being used uncritically for the evaluation of individual scholars, there are risks of intensification of overcompetitiveness and a widespread further demoralisation of much of the scholarly community. It will be necessary to institute some mechanisms for feedback, and that, at least in theory, is more likely to occur with an organisation funded directly by the taxpayer.

"Freedom of information" was a catchcry used by politicians and academics a few years ago. The concept was never rigorously explored. One form of implementation, the passage of Freedom of Information Acts or their equivalent, has had little effect except in the United States and Sweden. Despite the limited successes in penetrating unnecessary secrecy, the fact is that the control of many types of information is becoming increasingly concentrated into the hands of an oligopoly of information entrepreneurs and high-level bureaucrats. Nevertheless, the citation indices are a fascinating revelation in the flow of information. Like many inventions they can, at least temporarily, disturb existing power structures.

Ambivalence of Administrators to Bibliometrics: the Citation Indices Give Some Power to the Academic Proletariat — At Least Temporarily

We are unable to find information about how frequently academic or scientific administrators use the citation indices in deciding the fates of staff. Nicholas Wade[94] claims that citation analysis is used at some American universities as "part of the evidence for deciding cases of promotion and tenure". Wade writes too that "the National Science Foundation (NSF) is using the technique to assess its funding of chemistry departments and as a safety net to catch chemists who write bad grant proposals but are heavily cited". However, Wade's latter claim drew a denial from an administrator in the chemistry section of the NSF.[95]

Wade[96] also mentions that citation counts were used in evidence in a court case to show that a female biochemist at an eastern American university was discriminated against in being denied tenure. Her citation count was better than that of a male staff member who had been granted tenure.

What is of significance is that in 1979, years after the introduction of citation indices and after much discussion of their uses in scholarly journals, we met two prominent British academic administrators, both involved in "life-or-death" decisions over staff being considered for redundancy. Yet, neither of these academic administrators knew about the citation indices — and didn't seem to want to know.

Many of the books and articles cited by us earlier in this section provide evidence for a *partial* association between various measures of quality and the citation counts. If due allowance is made for the biases discussed briefly by us earlier, the method can be used on individuals. This does *not* mean that one individual who gets 10 citations a year is better than another individual who gets eight. In fact, the 95 per cent confidence interval for comparing the citation impact of two different individual papers is very large. Just the sampling variation is so great that the 95 per cent confidence intervals are three versus 10, 10 versus 20, or 99 versus 124 (see note 97). When one adds the counts over the years, and for a number of publications, the techniques become more discriminatory — providing that the interpretation of the numerical data is tempered with the caveats discussed here.

Probably the most important source of error is the ingroup effect, or the "public relations web" mentioned earlier. This tends to give dissidents and other outsiders a low score. One possibility is to use the method employed by Dennis Dieks and Hans Chang.[97] They measured "strangeness", the percentage of citations arising from outside the country of the cited. Dieks and Chang showed that this measure was different for different groups of Dutch workers publishing on various aspects of magnetic resonance spectroscopy and suggest that "strangeness"... "can be useful to identify 'incrowds' in science". Presumably, researchers in another country will be less influenced by local favouritism or disfavouritism.

Such an example may seem a striking exception to R. K. Merton's basic norms of science, notably *universalism* and *disinterestedness*.[98] However, we know of a number of instances where researchers refused to cite a dissident's papers, even though these papers had clear priority or described techniques which they had copied. One of the individuals actually confessed to us that he was afraid to cite these papers lest he offend a powerful senior academic.

The overriding importance of *status* in influencing citation practices is evident from the research performed by Richard Whitley, which also gives a useful suggestion for reducing bias from the "public relations web". Whitley studied the citation impact and practices of two groups of physiologists: HIPP physiologists, who were "high in power and prestige", versus LOPP physiologists, who were "no power and low prestige".[99] Although LOPP physiologists received only six citations in physiological journals, versus 156 citations for HIPP physiologists, this overwhelming difference was greatly reduced when citations in non-physiological journals were measured. In journals outside the field of physiology the LOPP physiologists received 165 citations compared with 500 citations for HIPP physiologists. In other words, the no-power-and-low-prestige physiologists were making a relatively greater impact outside their field than inside. The HIPP physiologists still had a greater total impact but their dominance inside the physiology journals (156/6 = 26) was weakened outside the physiology journals (500/165 = approximately 3). What is especially important is Whitley's observation that HIPP physiologists *never* cited LOPP physiologists. Rank rules!

Now, with the review of various biases, we can return to the question of why many scientific administrators are reluctant to use the citation indices in obtaining an *approximate* measure of quality. We believe that, despite the many biases inherent in the use of citation indices, these methods do take some power away from the administrator. *It is easy to get the desired result by seeking a couple of anonymous opinions about a researcher. Despite the many biases in the citation indices, it is far less easy to "tilt" the counts in an open publication, accessible to administrator and to dissident alike.*

Furthermore, the tables can be turned: the dissident can always check up on the scholarly impact of the administrator, or upon the impact of a complainant or his supporters.

In the course of defending dissidents we have noticed a wide range of behaviour of staff association officials. Some are helpful and ready to protest injustice, but others are not. It gradually became apparent that the more vigorous defenders of dissent had, on the whole, better records of publication and citation. These results go along with the literature reviewed towards the beginning of this chapter, the partial positive association between social dissidence and scholarly accomplishment. However, such results have an important implication in the defence of academic freedom: the selection of mediocre individuals for staff association positions can be a threat to intellectual freedom and fairness. Some staff association officials are well motivated and have sufficient scholarly status to be able to stand up against academics or administrators participating in a witch-hunt. Unfortunately, a few individuals use their election to staff association positions as a conduit into the corridors of administrative power.

We conclude the discussion of citation indices by returning to the couplet from Alexander Pope, used at the beginning of this section: "Index-learning" does now "hold the Eel of science by the Tail". When allowance is made for the biases, the citation indices can also help to catch the eel-climbing academic who advances by means distinct from scholarly recognition.

"When All Else Fails, Try Reading . . ."

We have passed from the deviousness of personal opinions to the elegance of bibliometrics. At this point of desperation in the discussion of methods for the evaluation of performance we draw attention to a relatively neglected means for the evaluation of both research and academic staff. We suggest that administrators should try reading the publications of staff about whom they must make decisions of hiring, firing, or promotion.

Many specialist papers can only be fully comprehended by similar specialists. But a senior academic administrator should at least be able to judge certain general qualities of books, review articles or more popular presentations. Clarity of expression, logical development and enthusiasm for the subject should be evident to the capable administrator. It may also sometimes be possible to judge originality, if the field is not too far removed from the administrator's experience. Such qualities are desirable in both researchers and teachers.

We have no data on the extent to which academic administrators, or staff members generally, actually examine the publications of those on whom they pass judgement. We hope that the following confession revealed in the interviewing of an academic administrator by Caplow and McGee does not reflect the norm:

Q: Are the men's publications read?

A: Oh, yes!

Q: By whom?

A: By the tenure members, at least.

Q: All of them?

A: Yes.

Q: Did you read them?

A: Yes.

Q: Did you read those of the man you finally hired?

A: Yes.

Q: What was the one which you remember best about?

A: Well... I didn't read it, exactly. I looked it over. It was in a good journal. Nothing trashy gets in there.

Q: What do you mean by you "looked it over"?

A: Well, I looked at it, looked at his references, read his abstract.

Q: Is that the way the rest of the committee handles the publications, do you think?

A: I think so, yes, they look them over.[100]

Evaluation of Teaching

> Powerful evaluation procedures can yield results from which the administrator may have no place to hide. Weak methods yield results that can be interpreted to his advantage. To many administrators weak methods seem better...
>
> [J. G. Stanley]
>
> Of course, not only teachers are opposed to evaluation. Administrators, who are wise in the way of politics and organizational behaviour, realize that reverberations may take place on the campus; once evaluation of students is echoed by evaluation of teachers it may not be long before the evaluation of the administration is called for.
>
> [G. L. Geis][101]

It is generally agreed that the evaluation of teaching is even more difficult than the evaluation of research. The topic of what makes a good teacher is one that is discussed frequently in books and papers, but with little agreement. A recent collection of review articles by the Canadian Association of University Teachers[102] provides a useful introduction to the literature on evaluation of teaching in higher education.

The lack of agreement about what constitutes good teaching is not surprising. Criticisms of teaching performance often do not allow for individual differences, both among teachers and among students. Certain styles of teaching suit some lecturers better than others. Conversely, some students prefer one type of teaching to another. A teacher who is considered dull and pedantic by one person might be considered well organised and carefully paced by another.

A further problem is that evaluation of teaching is coloured by one's own experiences, politics and value systems. One study of different teaching styles, assessed by different people, found that the same example could be judged "exemplary, trivial or unethical".[103]

Individual variation in learning ability cannot be ignored — especially in this time when more people are going on to higher education and when universities are encouraged (or forced) by politicians to reduce the ranks of the unemployed. Some students need "high arousal". They prefer exciting and enthusiastic speakers. When such "high arousal" students are inspired, they do excellent work. The slow-but-steady student, who may in the end reach a high standard of performance, usually does better with a more carefully paced and not-too-stimulating lecturer.

There is also now a pronounced generation gap between staff members in their opinions about what constitutes good (or bad) teaching. We believe that this divergence in opinion has probably arisen as a result of television and its influences on students. Thus, older and younger staff members often disagree about the means needed to encourage (or to coerce) students to study. For the last twenty years in Western countries, students entering university differ from their predecessors in their massive exposure to a novel means of communication. (They also differ from earlier students in being recruited from a broader social and ethnic base.)

The exposure to television, with viewing times of twenty to thirty hours per week, presents problems for teachers. Students have become conditioned, like Pavlov's dogs, to brief, pre-digested tidbits of knowledge, sugar-coated with a false lucidity. Concentration span has shrunk to the time for the average television commercial. It requires a determined personality on the part of the teacher to get students to break through the self-imposed barriers of mass mediocrity, indeed, just to get the students to read.

Television is also a one-way flow of information. If exposure to television makes students restless in ordinary lectures, it can also make them strangely passive recipients to knowledge. Some teachers, of course, prefer students who do not answer back. Other teachers deplore the intellectual deadness.

At the same time, television, for all its faults, has made the average student much more aware of what goes on in the world. Teachers who isolate their subject from the reality that students perceive run the risk of losing credibility, if not attention.

Whatever the differences in opinion about the value of television, few will deny that it hasn't changed attitudes about what constitutes good presentation. There are the rare examples of high quality lectures on television, combined with the visual presentation of examples from film or TV cameras in a way that would be beyond the funds available in any university.

Some qualities in a good teacher

In the course of interviewing candidates for academic jobs, it is common to ask them to give a seminar. This allows one to judge not only some of their research but, in particular, their ability to present material clearly. The results may not be representative of their day-to-day abilities at teaching a range of subjects, including, unavoidably, many for which the lecturer has not had first-hand research experience. Academic appointments more often go to specialists rather than generalists, but the specialist may have more difficulty in lecturing on the wide range of subjects necessary in teaching.

Other aspects of teaching performance need consideration besides lecturing. In the sciences a quite different approach is needed in organising effective laboratory practicals or field studies. Working with small groups, for example in tutorials, requires yet different talents.

Perhaps the most important quality in a teacher is being able to make effective contact with students, encouraging critical examination, both of their own efforts and the efforts of others. Some subtle factors are involved in balancing encouragement and discipline. A brilliant lecturer may be so personally arrogant that students are "turned off". We recall one lecturer who epitomised "radical chic" in his trendy dress and bizarre personal activities; but, he was the only staff member who could reach certain students and stimulate them to do excellent research. It takes all kinds to make an effective teaching world. By passing judgement too hastily, especially on nebulous concepts such as "personality", administrators reject many individuals who have that innate ability to make contact with certain students.

Some of the most effective teaching takes place in informal settings, often in casual conversation between staff member and student. Yet, in the evaluation of the performance of teachers, credit is rarely given to those who spend long hours helping, guiding, inspiring or just listening to students.

Thus, allowing for the wide range of opinions about what constitutes good teaching, the best procedure is to ask for a large number of independent assessments. Such evaluations should include some standard set of questions, such as: How well prepared is the lecturer? How well read is the lecturer? How effective is the lecturer in using visual aids? Does the lecturer show a willingness to answer students' questions? Does the lecturer respect the views of students, especially when they differ from his own? Does the lecturer inform students ahead of time how they will be graded? In grading examinations, does the lecturer give students a good idea of what errors they made?

Our list is not intended to be complete and, obviously, must be adjusted to the requirements of different subjects, in particular the relative emphasis on lectures, tutorials and laboratory practicals. The standard set of questions should be used in the evaluation of *all* staff, not just the individual being considered for promotion (or sacking). The results should be freely available so that there is no suspicion of abuse in secrecy, although, at least for ordinary situations, the privacy of individuals will need to be preserved. In other words, for routine matters, each lecturer would be given only his or her own results, the results for others being presented as a general numerical distribution. However, when dismissal charges are being prosecuted — charges alleging unsatisfactory teaching — then the right of the individual to a fair trial transcends the rights of other staff members to their privacy.

Student evaluation of staff teaching performance

So far we have said nothing about student opinion in assessing academic performance. This is anathema to many academics. The very suggestion that students should be permitted to express an opinion, let alone that it be given some weight in evaluation, has on more than one occasion contributed to a dissident being given a one-way ticket out of the academic profession. It is thus welcome to find the example of a Vice-Chancellor[104] who points out the salutary role of student opinion, including even the candid comments which occasionally occur in student newspapers or counter-calendars.

Student opinion is no better and no worse than any other opinion. It is sometimes coloured by the prevailing subculture of anti-intellectualism ("ockerism" in Australia) and by a general hostility to what it perceives as parental authority. But, beyond the generation gap and mediocrity cult, there is sometimes an intolerance for unfairness and a youthful ability to detect — and to puncture — pomposity. Students have often sampled more of a lecturer's efforts than his professional colleagues have; also, students will have an immediate perception of how those efforts compare with the efforts of other staff.

The sensible way to use student opinion is the same as for any other kind of opinion. Ask a number of questions (rather along the lines of those listed in the previous section). The questions should not be "loaded" and should allow for the multiple facets of good teaching. A reasonably large random sample of students should be taken. One should allow for the fact that a staff member who expects students to work hard will always receive complaints that the lectures are too difficult to understand or that the reading assignments are too long. After all, one is supposed to be assessing performance in a university, not a kindergarten.

What we find odious in some attempts to dismiss staff is where the administration presents the opinion of only one or two selected students. The administration never reveals how many students were sampled to get a few who would give the opinion they wanted. This is a standard of unsatisfactory sampling that those same academic administrators would not tolerate in their own research (or at least we would hope that they would not tolerate it).

The claim is sometimes made that students' opinions are influenced by the grade they receive.[105] That may not be as bad as it sounds. Since a student who likes a particular teacher often works harder in the subject, it is not surprising that there is a positive partial correlation between grades and student opinion. We know of no evidence that students are more likely to be subject to faulty judgement arising out of conflicts of interest than are academics. The answer to such problems is, as mentioned before, to ensure that the sampling procedure is valid and that there are data available for opinions about other staff besides the one being considered for administrative action.

It is unsettling to note that in several cases where dissidents were alleged to have had poor performance in teaching, their non-dissident colleagues, whose performance was much worse, were not considered for dismissal. When a staff member frequently misses his lectures, fails to turn in his grades on time, arrives at lectures late or unprepared, fails to keep up in the subjects about which he is lecturing, treats the students unfairly, or even indulges in what is euphemistically called "sexual harassment", why are not the dismissal statutes invoked? Such behaviour would appear to qualify as "gross and persistent neglect of duties". Senator Baden Teague, the Chairman of the Senate Standing Committee on Education and the Arts, investigating the tenure of academics, complained: "...one could wax quite loudly about this almost national scandal of teaching inadequacies amongst tertiary staff".[106] That may be a bit extreme, but a problem exists and it is not helped by administrators spending so much time and effort in trying to sack dissidents on inadequate charges when consistently poor teaching, plagiarism, false charges, and sexual exploitation of students by other staff go unpunished.

In summary, the essence of fair evaluation is to have a thorough comparison of the individual being considered for some administrative action (appointment, promotion, tenure, or dismissal) with the performance of appropriate peers. The limitations of different methods of evaluation must be recognised. If the administration really has a good case, it need not rely on procedural ploys. Nor need the administration rely on inadequate sampling, either of an individual's work or actions. As Marcia Lieberman[107] wrote in the quotation given at the beginning of this chapter: "Covert discrimination cannot exist without falsehood...Often the blame is placed on the victim".

References

1. Marcia R. Lieberman, "The most important thing for you to know", in: Gloria DeSole and L. Hoffmann (eds), *Rocking the Boat: Academic Women and Academic Processes* (New York: Modern Language Association of America, 1981), pp. 3–7, quotation from p. 5, with a change in paragraphing.
2. The establishment view of creativity is exemplified by Jonathan R. Cole and Stephen Cole, "The Ortega Hypothesis", *Science*, vol. 178, pp. 368–75 (1972): only a very small percentage of scientists have the "sacred spark" to make really significant contributions.

   However, the evidence marshalled by the Coles is capable of an alternative explanation. Since the research by the mathematical biologist A. J. Lotka in the 1920s, it has been known that the number of scientists producing n papers is roughly proportional to n^{-2}: see, for example, Derek de Solla Price, *Little Science, Big Science* (New York: Columbia University Press, 1963). In other words, the majority of scientists who publish, publish only one or two papers in their lifetime. *This resembles the life-table data for a species with a high rate of random mortality where the mortality rate is especially high at the early stages*, e.g., Raymond Pearl's "Type III" mortality: see C. J. Krebs, *Ecology* (New York: Harper & Row, 1972), p. 158. This is contrary to what we would expect, given the requirement for some minimal intelligence, motivation and perseverance in order to do successful science (or other forms of scholarship). Truncating the normal distribution of human abilities should still yield a much flatter mortality curve for publication. This result, the n^{-2} distribution, automatically biases

the results on citation which the Coles use as their main evidence for elitism. There are other limitations to the use of citation data and some of these are discussed later in this chapter.

This alternative explanation of data which the Coles choose to interpret in an elitist manner illustrates a problem that also underlies much of the evaluation of performance. Intellectuals subconsciously (and sometimes consciously) "push" their hypotheses, experimental designs, and interpretation of results in a direction determined by preconceived ideas (or outright prejudices) or the source of research funding. The reader is well advised to study: Brian Martin, *The Bias of Science* (Canberra: Society for Social Responsibility in Science, 1979).

3. R. D. Wright, "Prologue", in: W. H. C. Eddy, *Orr* (Brisbane: Jacaranda, 1961), pp. xiii–xvi.
4. Bertrand Russell, *Autobiography* (Boston: Little, Brown, 1957), vol. 2, years 1914–44, p. 332. Because of his pacifism in World War I, Bertrand Russell was dismissed from his lectureship at Trinity College, Cambridge: see Ronald W. Clark, *The Life of Bertrand Russell* (New York: Knopf, 1976), pp. 289–92. Because of his alleged sexual permissiveness, Bertrand Russell was prevented from taking up a position at the City College of New York: see John Dewey and Horace M. Kallen (eds), *The Bertrand Russell Case* (New York: Viking Press, 1941, and 1972 reprinting); and, Paul Edwards, "Appendix", in: Bertrand Russell, *Why I Am Not a Christian* (London: Allen & Unwin, 1957), pp. 207–59.
5. D. Dickson, "Firing spotlights plutonium exports", *Science*, 221, p. 245 (1983); S. Bhatia, "Electricity Board critic is dismissed", *The Observer* (London), 12 June 1983, p. 3; see also the anonymous editorial, "Don't fire the messenger", *New Scientist*, 23 June 1983, p. 838. A useful exchange of letters is: Trevor Broom, "CEGB answers back", *New Scientist*, 7 July 1983, p. 50; and, R. V. Hesketh, "CEGB fired", *New Scientist*, 28 July 1983, p. 29. Confirmation of Hesketh (and Ryle) has come from Tony Benn, formerly the Secretary of State for Energy in the United Kingdom. Benn is quoted as saying: "Every British nuclear power station becomes a nuclear bomb factory for the US" (*New Scientist*, 8 December 1983, p. 724).
6. Lieberman, op. cit.
7. T. Caplow and R. J. McGee, *The Academic Marketplace* (New York: Basic Books, 1958).
8. R. T. Spooner, "Will you give me a reference?", *Education*, 27 August 1976 issue, pp. 173–4.
9. Advertisement in *Science* 195, p. 1370 (1977).
10. J. P. Rushton, H. G. Murray and S. V. Paunonen, "Personality, research creativity, and teaching effectiveness in university professors", *Scientometrics* 5, pp. 93–116 (1983).
11. E. Anderson, *Plants, Man and Life* (Berkeley: University of California Press, 1967 reprinting; originally published in 1952).
12. H. Seyle, *From Dream to Discovery* (New York: McGraw-Hill, 1964).
13. A. Sibatani, "You carry out eukaryote experiments on shellfish selfish DNA: an essay on the vulgarization of molecular biology" in: *Science and Scientists: Essays by Biochemists, Biologists and Chemists* (Tokyo: Japan Scientific Societies Press, 1981); and, A. Sibatani "Molecular biology: a paradox, illusion and myth", *Trends in Biochemical Sciences*, June 1981 issue, pp. vi–ix.
14. Anne Roe, "A psychological study of eminent biologists", *Psychological Monographs*, vol. 65, no. 14, pp. 1–68 (1951); B. T. Eiduson, *Scientists: Their Psychological World* (New York: Basic Books, 1962); Liam Hudson, *Contrary Imaginations* (London: Methuen, 1966).

15. A. H. Halsey and M. A. Trow, *The British Academics* (London: Faber & Faber, 1971).
16. See the review and new data in: S. M. Lipset and G. M. Schaflander, *Passion and Politics: Student Activism in America* (Boston: Little, Brown, 1971) and E. C. Ladd, Jr. and S. M. Lipset, *The Divided Academy: Professors and Politics* (New York: McGraw-Hill for the Carnegie Commission on Higher Education, 1975).
17. Clark Kerr, *The Uses of the University* (New York: Harper and Row, 1966), p. 99.
18. H. Marcuse, "Repressive tolerance", in: R. P. Wolff, B. Moore, Jr., and H. Marcuse, *A Critique of Pure Tolerance* (Boston: Beacon Press, second printing, 1969), p. 119.
19. Anonymous, "Angry astronomers gag their critics", *New Scientist*, 17 February 1983, p. 424.
20. David Dickson, "Study of big science group hits raw nerve", *Science* 220, pp. 482–3, quoting p. 483 (1983).
21. J. Irvine and B. R. Martin, "Assessing basic research: the case of the Isaac Newton Telescope", *Social Studies of Science*, vol. 13, pp. 49–60 (1983); B. R. Martin and J. Irvine, "Assessing basic research: some partial indicators of scientific progress in radioastronomy", *Research Policy*, vol. 12, pp. 61–90 (1983).
22. With regard to the "fickleness in the reward system" and the failure of multiple awards to represent *independent* estimates of quality, see A. Carl Leopold, "The act of creation: creative processes in science", *BioScience* 28, pp. 436–40 (1978). Even the Nobel Prize has in its history a number of examples of serious controversy and belatedly recognised error; see, for example: Deborah Shapley, "Nobelists: Piccioni lawsuit raises questions about the 1959 prize", *Science* 176, pp. 1405–6 (1972); Wallace Cloud, "Winners and sinners", *The Sciences (New York Academy of Science)*, December 1973, pp. 16–21; Anonymous, "A Nobel scandal?", *Time* Magazine, 7 April 1975, p. 55; H. Inhaber and K. Przednowek, "Quality of research and Nobel prizes", *Social Studies of Science* 6, pp. 33–50 (1976); M. Benarie, "Nobel Prize rules", *Nature* 288, p. 8 (1980); Danah Zohar, "The science prizes: they get those wrong, too", *The Sunday Times* (London), 23 November 1980, p. 35.
23. Caplow and McGee, op. cit.
24. Brian Martin, "The scientific straightjacket", *Ecologist*, vol. 11, Jan./Feb. issue, pp. 33–43 (1981); Brian Martin, "The naked experts", *Ecologist*, vol. 12, July/Aug. issue, pp. 149–57 (1982).
25. For example, see the Senate Standing Committee on Education and the Arts, *Tenure of Academics* (Canberra: AGPS, September 1982), p. 83 and the relevant Hansard transcripts and submissions, especially Senator P. J. Giles, pp. 1731–5.
26. M. Oliphant, "The quality of Australian universities", *Vestes*, vol. 3, issue 2, pp. 45–9 (1961).
27. Select Committee to Study Governmental Operations With Respect to Intelligence Activities, US Senate, Book I: *Foreign and Military Intelligence*, and Book II: *Intelligence Activities and the Rights of Americans* (Washington, DC: US Government Printing Office, Report 94–755, 1976). By 1967 the CIA had sponsored more than a thousand books. For the names of a number of eminent academics whose research or publication were financed by the CIA see: John Marks, *The Search for the Manchurian Candidate: the CIA and Mind Control* (Allen Lane, Penguin Books, 1979).
28. W. M. O'Neil, *Advice to A.R.G.C. Applicants* (Canberra: Australian Research Grants Committee, 1972, mimeographed document). Subsequent versions of the advice to applicants retain much of O'Neil's mimeographed document but this telling passage has been deleted.
29. Robert Drewe, "How bureaucratic venom threatens your life", *Bulletin* (Sydney), 12 January 1982, pp. 18–24. The case history is included in the "Archives of Suppression" chapter.

30. Our comments are not to be interpreted as pardoning the absence of *some* output. An academic with approximately half of his time free to do research ought to publish about one paper a year, or a book or a major review article at less frequent intervals. Even when denied research funds, there is much to be done in the way of theoretical and review work. It is revealing of the standards of administration that some staff members are allowed to spend years without showing any evidence for productive work. Tenure is no excuse, for dismissal statutes would allow the sacking of staff who neglect their scholarly duties for long periods.
31. Caplow and McGee, op. cit., p. 153.
32. Donald Michie, "Peer review and the bureaucracy", *Times Higher Education Supplement*, 4 August 1978, p. 11.
33. D. Fifield, "Nature", *New Scientist*, 30 October 1969, pp. 230–2.
34. C. Manwell, "Peer review: a case history from the Australian Research Grants Committee", *Search*, 10, 81–6 (1979); I. I. Mitroff and D. E. Chubin, "Peer review at the NSF: a dialectical policy analysis", *Social Studies of Science* 9, pp. 199–232 (1979); D. P. Peters and S. J. Ceci, "Peer-review practices of psychological journals: the fate of published articles, submitted again", *Behavioral and Brain Sciences* 5, pp. 187–95 (1982); see also the comments by others in the same issue, pp. 196–255. The Peters and Ceci paper, together with the comments, have been published separately: S. Harnad (ed.), *Peer Commentary on Peer Review. A Case Study in Scientific Quality Control* (Cambridge University Press, 1982). See also the chapter "Prejudice in Granting Research Grants".
35. C. Manwell and C. M. A. Baker, "Honesty in science", *Search* 12, pp. 151–60 (1981); William Broad and Nicholas Wade, *Betrayers of the Truth* (New York: Simon & Schuster, 1982); Nicholas Wade, "What science can learn from science fraud", *New Scientist*, 28 July 1983, pp. 273–5; C. Ian Jackson and John W. Prados, "Honor in science", *American Scientist* 71, pp. 462–4 (1983).
36. Useful discussions about the differences in opinion among statisticians can be found in: S. Siegel, *Non-parametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill, 1956); L. Hogben, *Statistical Theory* (London: Allen and Unwin, 1957); and, R. R. Sokal and F. J. Rohlf, *Biometry* (San Francisco: W. H. Freeman, 1969).
37. S. M. Gore, I. G. Jones and E. C. Rytter, "Misuse of statistical methods: critical assessment of articles in BMJ from January to March 1976", *British Medical Journal*, 8 January 1977, pp. 85–7.
38. For an excellent discussion of the evils of secrecy see: Richard Davis, "Anonymity — the cancer of academia", *Education Research and Perspectives*, vol. 6, issue 2, pp. 3–11 (1979).
39. Lionel S. Lewis, *Scaling the Ivory Tower: Merit and Its Limits in Academic Careers* (Baltimore: Johns Hopkins Press, 1975). See also Caplow and McGee, op. cit. and Spooner, op. cit.
40. Lewis, op. cit.
41. See notes 34 and 35 above.
42. See note 39 above.
43. S. Cole, J. R. Cole and G. A. Simon, "Chance and consensus in peer review", *Science* 214, pp. 881–6 (1981). In their earlier work dealing with peer review the Coles took a strongly elitist position. Although they still maintain some of their earlier opinions, the data in this article argue strongly that peer review is characterised by a marked disagreement among assessors.
44. R. A. Fisher, "Has Mendel's work been rediscovered?", *Annals of Science* 1, pp. 115–37 (1936). R. A. Fisher's conclusions, that Mendel's data were deliberately trimmed (by Mendel or by his assistant), are disputed by several authorities: Sewall Wright, "Mendel's ratios" in: Curt Stern and E. R. Sherwood (eds), *The Origin of Genetics* (San Francisco: W. H. Freeman, 1966), pp. 173–5; Åke Gustafsson, "The life

of Johann Mendel — tragic or not?", *Hereditas* 62, pp. 239–58 (1969); B. L. Van der Waekden, "Mendel's experiments", *Centaurus* 12, pp. 275–88 (1968); Robert Scott Root-Bernstein, "Mendel and methodology", *History of Science* 21, pp. 275–95 (1983). The last-mentioned reference has missed Sir Gavin de Beer's suggestion (although citing the paper) that one of Mendel's assistants had an "inclination towards bibulousness" and worked "subconsciously in the direction of obtaining the results that were expected, by discarding doubtful material or bad specimens, and by making honest mistakes": G. de Beer, "Mendel, Darwin and Fisher", *Notes and Records of the Royal Society (London)*, vol. 19, pp. 192–226, quoting from p. 201 (1964). Robert Scott Root-Bernstein's assertion that "Mendel counted all the peas he grew" is too strong; even if Mendel or his assistants attempted to count every last pea, the point is reached where damaged material, accidental loss, or removal by pests, intervenes. Thus, although Root-Bernstein provides an interesting suggestion, which proves Mendel's honesty, the explanations by some earlier authors are not so easily ruled out. Perhaps the best evidence for Mendel's honesty is his willingness to publish research which did *not* conform to his ratios, the work on hawkweeds: see L. C. Dunn, "Mendel, his work and his place in history", *Proceedings of the American Philosophical Society* 109, pp. 189–98 (1965).

45. See note 34 above.
46. C. Manwell, "Dissident scientists: hard versus soft science", *Physics Bulletin* 29, pp. 267–8 (1978).
47. S. J. Ceci and D. P. Peters, "Peer review: a study of reliability", *Change*, September 1982 issue, pp. 44–8, quotations taken from pp. 46, 47 and footnote no. 3 on p. 48.
48. Lionel S. Lewis, "Publish or perish: some comments on a hyperbole", *Journal of Higher Education* 38, pp. 85–9 (1967). See also, L. S. Lewis, "Getting tenure: change and continuity", *Academe*, November 1980 issue, pp. 373–81. See also Peter Blunt, "Publish or perish or neither: what is happening in academia", *Vestes*, vol. 14, issue 1, pp. 62–4 (June 1976).
49. Lewis, *Journal of Higher Education*, op. cit. pp. 88–9.
50. W. J. Broad, "The publishing game: getting more for less: meet the Least Publishable Unit, one way of squeezing more papers out of a research project", *Science* 211, pp. 1137–9 (1981).
51. J. R. Cole and S. Cole, *Social Stratification in Science* (University of Chicago Press, 1973). While they have overstated their case, we would agree with the general conclusion that there is a weak association between quantity and quality. Such studies need to be corrected for *visibility* (a person who publishes lots of papers is more likely to get his work seen in the random browsing of journals by readers) and for *opportunity* (the very fact that the reward system favours quantity means that prolific publishers will get more grants and more students to do more, and sometimes better, research). For a discussion of other biases in the work of the Coles see C. Manwell and C. M. A. Baker, "Reform peer review: the Peters and Ceci study in the context of other current studies of scientific evaluation", *Behavioral and Brain Sciences* 5, pp. 221–5 (1982); and, Stephen P. Turner and Daryl E. Chubin, "Chance and eminence in science: Ecclesiastes II", *Social Science Information* 18, pp. 437–49 (1979).
52. Eugene Garfield, *Citation Indexing: Its Theory and Application in Science, Technology and Humanities* (New York: Wiley, 1979).
53. Harriet Zuckerman and Robert K. Merton, "Patterns of evaluation in science: institutionalisation, structure and functions of the referee system", *Minerva* 9, pp. 66–100, see especially Table 1, p. 76 (1971).
54. A. R. Kraft, J. A. Collins, L. C. Carey and D. B. Skinner, "Art and logic in scientific communications: abstracts, presentations, and manuscripts", *Journal of Surgical Research* 26, pp. 591–604 (1979); L. Goldman and A. Loscalzo, "Fate of cardiology research

originally published in abstract form'', *New England Journal of Medicine* 303, pp. 255–9 (1980).

55. Garfield, op. cit.
56. A. G. Heffner, ''Authorship recognition of subordinates in collaborative research'', *Social Studies of Science* 9, pp. 377–84 (1979).
57. Harriet Zuckerman, ''Patterns of name ordering among authors of scientific papers: a study of social symbolism and its ambiguity'', *American Journal of Sociology* 74, pp. 276–91 (1968).
58. Broad, op. cit.
59. L. H. T. West, T. Hore and P. K. Boon, ''Publication rates and research productivity'', *Vestes*, vol. 23, issue 2, pp. 32–7 (1980).
60. Lieberman, op. cit.
61. Stephen Fretwell, ''The impact of Robert MacArthur on ecology'', *Annual Review of Ecology and Systematics* 6, pp. 1–13 (1975).
62. Phillip M. Boffey, *The Brain Bank of America: An Inquiry into the Politics of Science* (New York: McGraw-Hill, 1975); Claude E. Barfield, ''Science Report/National Academy of Sciences tackles sensitive policy questions'', *National Journal* 3, pp. 101–12 (1971). That protégé favouritism has resulted in some substandard papers getting into the Proceedings of the US National Academy of Sciences is only part of the problem. There is also the matter of poor quality reports, often biased in favour of big business or government; see, for example, Nicholas Wade, ''Letter from Washington: credibility counts'', *Trends in the Biochemical Sciences*, September 1980 issue, p. xiii. Some similar problems exist in the Australian Academy of Science; see the penetrating review by Ann Moyal, ''The Australian Academy of Science: the anatomy of a scientific elite'', parts I and II, *Search* 11, pp. 231–9 and 281–8 (1980). Ann Moyal reports (p. 283): ''In interviews with the author, many Fellows sharply criticised Academy Reports . . .''
63. See note 35 above.
64. W. A. Hendrickson, R. E. Strandberg, A. A. Liljas, L. M. Amzel, and E. E. Lattman, ''True identity of a diffraction pattern attributed to valyl t-RNA'' (correspondence), *Nature* 303, p. 195 (1983); see also comments in the same issue, pp. 196–7.
65. Eugene Garfield, ''SCI Journal Citation Reports: a bibliometric analysis of science journals in the ISI* data base'', *Science Citation Index*, vol. 14 (1980).
66. Robert K. Adair and George L. Trigg, ''Editorial: should the character of Physical Review Letters be changed?'', *Physical Review Letters* 43, pp. 1969–74 (1979).
67. Lynn Margulis, ''Peer review attacked'' (letter), *The Sciences (New York Academy of Sciences)*, vol. 17, January/February issue, pp. 5, 31 (1977).
68. Alexander Pope, *Dunciad*, A, Book I, lines 233–4 (1729); version edited by James Sutherland (London: Methuen, 1943), p. 90.
69. Garfield, op. cit.
70. M. Oromaner, ''Professional age and the reception of sociological publications: a test of the Zuckerman–Merton hypothesis'', *Social Studies of Science* 7, pp. 381–8 (1977).
71. S. Cole, J. R. Cole and L. Dietrich, ''Measuring the cognitive state of scientific disciplines'', in: Y. Elkana, J. Lederberg, R. K. Merton, A. Thackray and H. Zuckerman (eds), *Towards a Metric of Science* (New York: Wiley, 1978), pp. 209–51, especially p. 223.
72. S. M. Lawani, ''Citation analysis and the quality of scientific productivity'', *BioScience* 27, pp. 26–31 (1977).
73. G. N. Gilbert, ''Referencing as persuasion'', *Social Studies of Science* 7, pp. 113–22 (1977).
74. Rodger Mitchell, ''Scaling in ecology'' (letter), *Science* 184, p. 1131 (1974).
75. Leigh Van Valen, ''Note on the 'Public relations web' '', *Evolutionary Theory*, vol. 1, issue 4, p. 106 (1975).

76. Timothy Lenoir, ''Quantitative foundations for the sociology of science: on linking blockmodeling with co-citation analysis'', *Social Studies of Science* 9, pp. 455–80 (1979), p. 478.
77. Nicholas Wade, ''Guillemin and Schally: a race spurred by rivalry'', *Science* 200, pp. 510–13 (1978), pp. 510–12; see also: N. Wade, *The Nobel Duel* (New York: Anchor/Doubleday, Garden City, 1981).
78. Derek J. de Solla Price, *Little Science, Big Science*, cited in reference 2, quoting from p. 78.
79. Gustav Nossal, ''Information flow within scientific peer groups'', in *59th Annual Review: Director's Report* (Victoria: Walter and Eliza Hall Institute of Medical Research, 1978), p. 8.
80. Loren Eisley, *Darwin's Century* (New York: Doubleday, 1958); and, L. Eisley, *Darwin and the Mysterious Mr. X* (New York: Dutton, 1979).
81. Eugene Garfield, ''Citation indexing for studying science'', *Nature* 227, pp. 669–71 (1970). See also the references cited in note 44 above.
82. T. S. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, 1970, second edition).
83. Erwin Chargaff, ''Triviality in science: a brief meditation on fashions'', *Perspectives in Biology and Medicine* 19, pp. 324–35 (1976), p. 324. Chargaff's writings provide entry into many controversial corners of science. See, for example: E. Chargaff, ''Building the tower of babble'', *Nature* 248, pp. 776–9 (1974); and, E. Chargaff, *Heraclitean Fire: Sketches from a Life before Nature* (New York: Rockefeller University Press, 1978).
84. R. Over and S. Smallman, ''Citation idiosyncrasies'', *Nature* 228, p. 1357 (1970).
85. J. D. Frame, F. Narin and M. P. Carpenter, ''The distribution of world science'', *Social Studies of Science* 7, pp. 501–16 (1977).
86. Jon Wiener, ''Footnote — or perish'', *Dissent* (New York), fall 1974 issue, pp. 588–92.
87. E. F. Hartree, ''Reprint distribution'', *Nature* 242, p. 485 (1973).
88. William J. Broad, ''Librarian turned entrepreneur makes millions off mere footnotes'', *Science* 202, p. 857 (1978).
89. ibid.
90. Wiener, op. cit.
91. Broad, op. cit.
92. Garfield, op. cit., p. 671.
93. Broad, op. cit.
94. Nicholas Wade, ''Citation analysis: a new tool for science administrators'', *Science* 188, pp. 429–32 (1975), p. 429.
95. O. W. Adams, ''NSF and citation analysis'', *Science* 189, p. 86 (1975).
96. Wade, op. cit.
97. D. Dieks and H. Chang, ''Differences in impact of scientific publications: some indices derived from a citation analysis'', *Social Studies of Science* 6, pp. 247–67 (1976), especially p. 257.
98. Robert K. Merton claimed that the behaviour of scientists followed four *norms: Universalism:* ''The acceptance or rejection of claims entering the lists of science is not to depend on the personal or social attributes of their protagonist . . . The imperative of universalism is rooted deep in the impersonal character of science'' (p. 607). ''*Communism*'' (sometimes called *communality*): the complete sharing of information (basically through publication in the open literature) — ''Secrecy is the antithesis of this norm; full and open communication its enactment'' (p. 611). *Disinterestedness*: self-interest is suppressed — the ''virtual absence of fraud in the annals of science . . . There is competition in the realm of science . . . and under competitive conditions there may well be generated incentives for eclipsing rivals by illicit means. But such impulses can find scant opportunity for expression in the field of scientific research'' (p. 613). *Organized*

Scepticism: "The suspension of judgment until 'the facts are at hand' and the detached scrutiny of beliefs in terms of empirical and logical criteria..." (p. 614). The quotations are from Robert K. Merton, *Social Theory and Social Structure* (New York: Free Press, 1968 enlarged edition). There appears to have been considerable confusion between public and private ideologies, between ideals, expressed attitudes, and actual behaviour. Merton, as a pillar of the American "eastern establishment" in the sociology of science, has had his norms accepted by many in this field; however, there have been some important challenges or modifications to Merton's norms. See, for example, S. B. Barnes and R. G. A. Dolby, "The scientific ethos: a deviant viewpoint", *Archiv europ. sociologie* 11, pp. 3–24 (1970); Robert A. Rothman, "A dissenting view of the scientific ethos", *British Journal of Sociology* 23, pp. 102–8 (1972); Ian I. Mitroff, "Norms and counter-norms in a select group of Apollo moon scientists: a case study of the ambivalence of scientists", *American Sociological Review* 39, pp. 579–95 (1974); and, Michael Mulkay, "Interpretation and the use of rules: the case of the norms of science", *Transactions of the New York Academy of Sciences*, series II, vol. 39, pp. 111–25 (1980). Basically, Merton's norms are largely ideals. In practice scientists do not always behave ideally. For example, demands on time and the limits to human information processing force many scientists to judge some research on the basis of *who* published it, i.e., ascription. This violates the norm of *universalism*. The "anthropological" style of research conducted by Latour and Woolgar, who observed the behaviour of scientists in a highly competitive laboratory, revealed that the initial reception of nearly all ideas and facts was determined by *who* originated them, a striking refutation of Merton's universalism; indeed, there appeared to be a pathological obsession about the personality characteristics and other attributes of individuals used in evaluating performance. See: Bruno Latour and Steve Woolgar, *Laboratory Life: the Social Construction of Scientific Facts* (Beverly Hills: Sage, 1979).

99. R. D. Whitley, "Communication nets in science: status and citation patterns in animal physiology", *Sociological Review* 17, pp. 219–33 (1969).
100. Caplow and McGee, op. cit., p. 127.
101. C. K. Knapper, G. L. Geis, C. E. Pascal and B. M. Shore (eds), *If Teaching Is Important: the Evaluation of Instruction in Higher Education* (Clarke, Irwin and Co. in association with the Canadian Association of University Teachers, 1977), p. 19.
102. ibid.
103. N. L. Smith, "Sources of values influencing education evaluation", *Studies in Educational Evaluation* 6, pp. 101–18 (1980), p. 101.
104. D. R. Stranks, testimony given before the Senate Standing Committee on Education and the Arts, Adelaide, 11 February 1982, p. 1727.
105. J. E. Dolin, testimony given before the Senate Standing Committee on Education and the Arts, Adelaide, 11 February 1982, p. 1791.
106. Baden Teague, comments recorded in the proofs for the Hansard record of the Senate Standing Committee meeting in Melbourne, 9 February 1982, p. 1389.
107. Lieberman, op. cit.